LAND INEQUALITY IN INDIA: NATURE, HISTORY, AND MARKETS

NITIN KUMAR BHARTI
DAVID BLAKESLEE
SAMREEN MALIK

https://wid.world/document/land-inequality-in-india-nature-history-and-markets-world-inequality-lab-working-paper-2026-09/

WORKING PAPER N°2026/09

APRIL 2026

WORLD INEQUALITY LAB

Land Inequality in India: Nature, History, and Markets

Nitin Kumar Bharti[*] David Blakeslee[†] Samreen Malik[‡]

March 1, 2026

Abstract

Land is the primary productive asset in agrarian economies. However, the origin of inequities in land holdings is poorly understood. Using data from 270,000 Indian villages, we disentangle the contributions of agricultural, market, and historical factors using a variety of methodological approaches. British rule and landlord tenure increase inequality—which is confirmed by a border discontinuity analysis—as does the presence of historically marginalized populations. Agricultural suitability increases inequality by expanding large holdings, though its impact is attenuated by structural transformation. Public goods provision has a concave relationship to land inequality, with a decline in villages dominated by one landlord.

Keywords: Land Inequality, Agriculture, Irrigation, Land tenure. *JEL codes:* Q15, O15, N55, D31, O13.

1 Introduction

Land is a pivotal economic, political, and social asset in agrarian societies. However, in many rural settings, a small number of households control a large share of agricultural land, while a large population remains landless or operates marginal plots (FAO, 2014; Lowder et al., 2021). Three broad forces are often invoked to explain these patterns, though they are typically studied in isolation. First, geography and agricultural productivity shape the returns to land, potentially encouraging consolidation where there are returns to scale, or where land is more valuable. Second, historical institutions, such as colonial governance, have left durable imprints on ownership

[*]Email: nitin.bharti@uwa.edu.au. University of Western Australia.
[†]Email: david.blakeslee@nyu.edu. New York University Abu Dhabi.
[‡]Email: samreen.malik@nyu.edu. New York University Abu Dhabi.

We thank Dimple Chopde and Ravi for excellent research assistance. We are grateful to Pulak Ghosh for giving access to the land ownership data, without which this analysis would not have been possible. We would like to thank Robert Allen, Guilhem Cassan, Oliver Vanden Eynde and Thomas Piketty for their comments and support.

How to cite this working paper: Bharti, N., Blakeslee, D., Malik, S., Land Inequality in India: Nature, History, and Markets, World Inequality Lab Working Paper 2026/05

patterns. Third, market-access and economic integration of rural regions alter the feasibility of land reallocation, as well as the profitability of farming relative to non-agricultural economic activities.[1] Whether these forces operate as substitutes or complements, and whether structural transformation and market integration mitigate or reinforce historically rooted inequalities, remain open empirical questions.

We address these questions using a comprehensive national census of land ownership, covering 650 million individuals from 270,000 villages across India—to the best of our knowledge, the first such dataset for a large developing country—which we integrate with detailed geospatial datasets on demographics, economic conditions, institutional details, and agro-ecological characteristics. This dataset allows us to build village-level land inequality measures and conduct a detailed and precise analysis of land inequality determinants, simultaneously disentangling the contribution of numerous factors; while also overcoming the challenges of ecological inferences arising in existing research, which generally focuses on country-level variation.

Using a rich set of land ownership inequality measures, we study the influence of three broad factors influencing land inequality: (1) agricultural suitability, or "first nature"; (2) historic legacies; and (3) market-access. Agricultural suitability are those geographic and agro-climatic characteristics which influence agricultural productivity; historic factors include legacies of British imperial rule, as well as caste structure; and market related factors- accounts for proximity to towns and transport networks. The Indian context is particularly well-suited to such an analysis due to large spatial variation in agro-climatic conditions, varied institutional legacies due to its colonial history, and substantial variation in economic conditions. Moreover, the diversity of land inequality levels across Indian states is almost as large as that between countries at the world level (Appendix Figure A1).

Indian villages feature extreme land inequality. The mean village-level Gini (on 0-100 scale) is 71.1, while the Gini among landholding households (in short "landowner Gini") is a more modest 45.9. 46% of households are landless. Among landed households, the average size of a landholding is 6.2 hectares. However, 28.9% of land is held by households with 0–1 hectares, and 48.6% by households with 1–2 hectares. The mean share of land held by the largest landholder in the village is 12.4%, while in 3.8% of villages the largest landholder owns more than 50% of the land.[2]

[1]The first factor, commonly referred to as "factor endowments" (Engerman and Sokoloff, 2000) or "first nature" (Easterly, 2007; Henderson et al., 2018), has generally only been studied in the context of suitability to plantation agriculture, while ignoring more general relationships between productivity and inequality. History plays a crucial role in today's development (Nunn (2020), and historical colonial institutions have been widely studied (Frankema, 2010; Banerjee and Iyer, 2005). The literature has studied a variety of different modes of transportation for market access such as highways (Allen and Atkin, 2022; Asturias et al., 2019), railroads (Donaldson and Hornbeck, 2016; Donaldson, 2018; Hornbeck and Rotemberg, 2024; Fenske et al., 2023), and proximity to towns (Fafchamps and Shilpi, 2005). Generally, market-access is found to have an overall positive outcomes in labor market conditions, economic activities, and farmers' real incomes, to name a few.

[2]Though the village-level dataset used in this paper is cross-sectional, district-level land inequality estimates

To understand the relative importance of the three factors, we perform a Shapley decomposition that breaks down the explained variation of the dependent variable across sets of independent variables.[3] This analysis reveals that agricultural suitability and historic factors are most predictive of inequality—7.0% and 7.8% of explained variation, respectively—followed by market-related factors (3.5%), which jointly explain 18.3% of variation. In order to better understand the role of individual set of factors in driving inequality, we estimate a variety of specifications, including semi-parametric methods, as well as spatial regression discontinuities.

We find that areas with higher agricultural suitability have higher land inequality. The relationship is strongly positive and linear upto the 60th percentile of predicted agricultural productivity, and levels off beyond that. To the best of our knowledge, this finding is unique in the literature; and represents an important insight distinct from that on the relationship between plantation agriculture and inequality (Engerman and Sokoloff, 2000; Easterly, 2007; Frankema, 2010). We show that this is driven by a substantial increase in the share of land held by large landlords (owning > 10 hectares), which comes at the expense of small farms, the owners of which become landless. We complement this analysis by looking at the impact of (exogenously) being within an irrigation scheme, or "command area"—which is associated with increases in agricultural productivity (Blakeslee et al., 2023; Asher et al., 2024)—on inequality. We find that the land inequality (Gini coefficient) increases by close to 1 percentage point due to being within a command area.

The impact of historical institutions on land distribution is strongly persistent. Villages in "princely states" exhibit lower levels of land inequality amongst all households compared to directly ruled British areas, on average.[4] This is driven primarily by a lower share of landless households, as there is no significant impact when examining the landowner Gini alone. In contrast, *zamindari* (landlord) areas clearly show higher levels of land inequality, driven by a marked reduction in the share of small farmers and a notable presence of dominant landlords. These results are buttressed by a rigorous border regression discontinuity design analysis. This analysis reveals that land inequality is approximately 2–3 pp lower within princely states and 3–4 pp higher in places featuring the *zamindari* system, complementing the findings of Iyer (2010) and Banerjee and Iyer (2005) on the development impacts of these institutions.[5] We also find that places with

from all-India representative surveys indicate persistence in land inequality over multiple decades (Banerjee and Somanathan, 2007).

[3]The use of the Shapley decomposition is motivated by the approach of Henderson et al. (2018), adapted to our framework.

[4]"Princely states" or *Native states* areas were governed by indigenous royal households, who exercised a supervised autonomy over many policy matters. These were slightly less than half of the total area and one-quarter population of total India in 1911 (Iyer, 2010).

[5]Banerjee and Iyer (2005) look at inequality impacts (among other outcomes), while Iyer (2010) does not study inequality impacts directly. Both use a district-level analysis. Our newly created princely states shapefiles, in conjunction with village-level land inequality data, allow us to perform a border regression discontinuity, departing

a higher Scheduled Caste (SC) population share have substantially higher levels of inequality, which is driven entirely by landlessness.

The market access-variables are also predictive of land inequality. The villages which are in closer vicinity to towns, large highways, and railroad stations have greater inequality.[6] Moreover, we see that inequality is elevated up to further distances from towns (10kms) than roads and railway stations (2.5kms). Additionally, villages with a bank or an agricultural market (*mandi*) have higher inequality, but there is no relationship in nearby villages.

The extent of structural transformation—i.e., the growth of the non-agricultural sector—has important impacts on the influence of agriculture on land inequality. Where the economy has shifted away from agricultural production, the role of agricultural suitability is entirely eliminated. Similarly, being within the vicinity of a town, or possessing an agricultural market, leads to a 30–40% reduction in the role of agricultural suitability. This finding implies that economic modernization and market integration fundamentally change the patterns of land inequality, though it is unclear what the impacts are for overall inequality, as erstwhile agriculture-based inequities may simply perpetuate themselves in the modern economy. In contrast, structural transformation and market-access have no impact on historic explanatory variables, indicating that the latter may cause forms of inequality less amenable to amelioration through market mechanisms.

We conclude our analysis by examining the consequences of land inequality for public goods provision. A substantial literature has argued that inequality reduces the provision of public goods, and that this in part explains the detrimental effects of inequality on long-run development (Engerman and Sokoloff, 2000; Easterly, 2007; Galor et al., 2009). We generally find a positive and concave relationship between inequality and public goods, which holds even when controlling for a broad array of broader development outcomes; as well as a positive relationship between landlessness and public goods. This suggests that, at the local level, some degree of inequality is potentially beneficial (as found in Banerjee and Somanathan, 2007), perhaps due to the ability of wealthier landowners to more effectively lobby for state resources; and that the presence of a large, socio-economically deprived population also leads to an inflow of resources. However, at higher levels of inequality this benefit declines; and, in fact, the presence of a single dominant landlord is associated with a decline in public goods, even when conditioning on overall levels of inequality.

Relatively little research has been done on the drivers of land inequality. One notable exception is the literature dealing with the role of colonial legacies, including both the use of plantation agriculture (Engerman and Sokoloff, 2000; Easterly, 2007; Frankema, 2010), as well as a variety

from the instrumental variable (IV) used in Banerjee and Iyer (2005) and Iyer (2010).

[6] Highways account for around 75% of total freight movement (Chaudhury, 2005), while railways also play an important role (Donaldson, 2018) in the Indian context.

of institutional innovations introduced by colonial authorities (Banerjee and Iyer, 2005; Ratnoo, 2024). Other factors invoked to explain land inequality include: population density (Alfani, 2022; Bardhan et al., 2014); agricultural technology (Braverman and Stiglitz, 1989; Bequet, 2024); and trade integration (Berg et al., 2023; Rao et al., 2022). A related literature has studied the effects of land reforms, which have been undertaken in many countries in order to reduce historic inequities (Binswanger-Mkhize et al., 2009; Frankema, 2005).

This paper makes a novel contribution in disentangling the factors driving land inequality. The substantial role played by agricultural suitability is, to the best of our knowledge, unique within the literature; while the acute levels of inequality arising from caste and landlord tenure reveal the persistence of inequities arising from history and social structure. The fact that structural transformation and market integration significantly attenuate the role of agricultural suitability, but not of history, reveals the limits of market mechanisms in rectifying historic inequities.

The remainder of the paper is structured as follows. Section 2 reviews the broader literature on land inequality and provides a focused discussion of its determinants in the Indian context. Section 3 describes our data sources and presents key descriptive statistics. In Section 4, we present our empirical results. Section 5 discusses the implications of land inequality for the provision of public goods, providing insights into its broader impacts on social welfare. Section 6 concludes.

2 Understanding Land Inequality

2.1 Drivers of Land Inequality

Farms are generally smaller in developing countries than in wealthier countries, with potentially detrimental impacts on productivity (Adamopoulos and Restuccia, 2014). Moreover, due to population pressures, farm size has been decreasing in recent years in low- and lower-middle-income countries, while increasing in upper-middle-income countries (Lowder et al., 2016).

Substantial differences in land inequality exist across developing countries, with important regional patterns (Bauluz et al., 2025; Frankema, 2005). Latin America has long been notorious for its high levels of inequality, as has South Asia to a lesser extent. In contrast, East Asia has generally featured lower levels of inequality, largely due to the successful implementation of land reforms. Sub-Saharan Africa has intermediate levels of inequality, though the situation is complicated by the continuing prevalence of communally held land.[7]

Easterly (2007) broadly classifies the drivers of inequality as being either "structural" or "market"-based. The former consist of "historical events such as conquest, colonization, slav-

[7] Jayne et al. (2016) studies the evolution of farm distribution in four African countries—Ghana, Kenya, Tanzania, and Zambia—to highlight the rise in medium-size holdings by urban-based professionals and influential rural people.

ery, and land distribution by the state or colonial power," whereas the latter are the outcomes of free market interactions and differences in farmer productivity. Additional factors have also been cited in the literature, including technological change, population density, and others, to which we return below.

The most extensive literature focuses on the fore-mentioned "structural" factors, particularly colonial legacies and land reforms. For example, the high levels of inequality in Latin America are generally attributed to the use of plantation agriculture by colonial powers in the presence of land well-suited to this type of cultivation (Engerman and Sokoloff, 2000). Frankema (2010) provides a comprehensive empirical analysis of country-level factors, seeking to disentangle the respective roles played by historical, geographic, and institutional factors using a cross-country analysis; and finds that land suitability for food crops is associated with lower inequality, while both British and Spanish rule are associated with higher inequality than rule by other colonial powers. Others have noted that within the post-colonial context of Latin America, the relationship between factor endowments and land inequality depends crucially on political factors (Nugent and Robinson, 2010).

The impact of land reforms has also been extensively studied, and yields conflicting findings.[8] The lower levels of inequality in East Asia are due to the successful implementation of land reforms implemented in the post-WWII era (İşcan, 2018), whereas persisently high inequality in Latin America and South Asia may be attributed to the failure of land reform (Besley and Burgess, 2000; de Janvry and Sadoulet, 1989). Even where land reforms have succeeded, benefits have not always redounded to the poorest elements of rural society: agrarian land reforms in Egypt in 1952 led to a reduction in land area inequality, but generally benefited only the middle classes (Moshrif, 2025); while land reforms in Brazil in the 1990s did not benefit the bottom end, and caused no significant change in landlessness (Assunção, 2008). Furthermore, successful land reforms have at time exposed the tension between equity and efficiency, as in the case of the Philippines, where the imposition of a land ceiling in 1988 led to a decline in land concentration, but also declines in productivity (Adamopoulos and Restuccia, 2020).

Other studies on land reforms have looked at programs that sought to improve the functioning of markets for agricultural land. Boberg-Fazlić et al. (2022) find that pro-market agrarian reforms in 18th-century Denmark led to greater productivity gains, though at the expense of rising land inequality, with the losers emigrating to the US.[9] De Janvry et al. (2015) finds that the abolition of the Mexican *ejido* system led to out-migration and land consolidation. Markevich and Zhuravskaya (2018) finds agricultural productivity gains from the abolition of serfdom in Russia,

[8] See Deininger et al. (2003) and Binswanger-Mkhize et al. (2009) for an overview.

[9] Additional historical studies from 18th-century land reforms include Galli and Rönnbäck (2021) in Sierra Leone, Finley et al. (2021) in France.

and losses from subsequent collectivization, though the authors are not able to look at impacts on inequality itself.

Another approach to explaining inequality comes from Braverman and Stiglitz (1989), who present a theoretical model showing that agricultural productivity-enhancing technologies can lead to an increase in land inequality if they differentially benefit the wealthy. The mechanism these authors cite is that of differential returns to new agricultural technologies arising from the necessity of complementary inputs, though anything giving rise to differential gains will have a similar effect. Evidence for such a mechanism was found recently by Bequet (2024), who showed increases in inequality in the Philippines after the introduction of GM corn. Recent work on the impact of highway improvements in India (Berg et al., 2023) as well as town proximity (Rao et al., 2022) have invoked related mechanisms to explain increases in inequality.

Work on the ancient origins of inequality provides additional evidence for the impact of technological progress. One prominent hypothesis argues that the introduction of the plow led to an increase in inequality due to the costliness of procuring and operating a plow (Bogaard et al., 2017; Bowles and Fochesato, 2024).

Population density has also been shown to play a role in determining inequality levels. The Black Death in the 14th century in Europe caused massive population decline, leading to land abundance and a reduction in land inequality (Alfani, 2022). However, not all exogenous population changes had a similar impact: later pandemics in Europe had only a marginal impact on land inequality despite causing significant declines in population (Alfani, 2010); while a population collapse in Mexico in the 16th century actually led to an increase in land concentration (Sellars and Alix-Garcia, 2018). Bogaard et al. (2017) show that places in the ancient world with greater population density have higher levels of inequality. However, population density is generally highly endogenous to agricultural suitability and trade connectivity (Henderson et al., 2018), making it challenging to disentangle its impact in the absence of population shocks.

Drawing on the typology of Easterly (2007), and incorporating the insights of other research, we study three broad drivers of inequality: (1) "first nature," which are those geographic attributes which contribute to agricultural potential; (2) "structural" factors, which we label as "historic"; and (3) market-access, which is related to, but slightly distinct from, the "market factors" invoked by Easterly (2007). The first category has hitherto been generally neglected in the literature; however, as we show subsequently, places in India with higher agricultural suitability are systematically more unequal. It should be noted that we do not study land reforms here, as these vary at the state level, with numerous distinct reforms in each state; and are therefore impossible to disentangle from other state-level factors shaping inequality in our sample.

The role of "first nature" in driving inequality is likely related, at least in part, to both the technology-related mechanism and the population-density channel. For example, the impact of agricultural suitability may derive in part from differential returns to the types of agriculture permitted by the agro-ecological environment. Alternatively, greater agricultural suitability may drive higher population density, and thereby contribute to inequality, perhaps through the land fragmentation channel highlighted by Bardhan et al. (2014). Even if these factors are at work in the impact of land quality on inequality, we regard them as some of the causal channels through which agricultural suitability impacts land inequality.

Complicating any analysis of the determinants of land inequality are both a general paucity of data, as well as ambiguity in the appropriate measures. While income inequality has in many cases been used as a proxy for land inequality due to data limitations (e.g., Easterly, 2001, 2007), Deininger and Squire (1998) show that income inequality is only imperfectly related to land inequality, with a Gini correlation of 0.39. Moreover, how one accounts for landlessness has important implications for inequality measures (Erickson, 2004), as does accounting for both area and *quality* of land (Bauluz et al., 2025).

2.2 Drivers of Land Inequality, India

Land inequality shows substantial spatial variation across Indian states (Figure 1). While some states feature levels of inequality that would rank relatively low globally, comparable to those of Sub-Saharan African countries, others would be as high as those of the most unequal countries in the world (Appendix Figure A1).

A variety of factors are present in India that may help explain some of these patterns. For example, during the colonial era, the British government introduced different land revenue systems in different parts of India with the primary objective of increasing land revenue and rents. In some areas, this entailed the investing of erstwhile tax farmers (*zamindars*) with formal ownerhip rights, leading to the creation of an enduring landlord class; which led to lower levels of economic and political development, and greater poverty and inequality (Banerjee and Iyer, 2005; Misra, 2019; Verghese and Teitelbaum, 2019; Caum-Julio, 2024).

Another important historical factor impacting development patterns is the presence of the so-called “princely states,” which were polities governed by indigenous royal households, who exercised a supervised autonomy over many policy matters. Researchers have shown that these areas are somewhat more developed than British-ruled areas (Iyer, 2010; Jha and Talathi, 2024; Iyer and Weir, 2025).[10] However, there is some evidence that caste-based inequities (discussed further below) are more severe within princely states: using a spatial RDD design, Varun (2024) finds that caste inequities are greater within the princely state of Hyderabad than in adjacent, non-princely state areas.

[10] Roy (2014) argues that some of the impacts of these historic and institutional features may derive from geographic characteristics and their interaction with post-Independence economic trends.

The fore-mentioned studies exploit state- or district-level variation in their analyses.[11] However, this gives rise to problems of ecological inference that may be particularly pronounced due to strong spatial correlations in historic institutions (Lee, 2023). More recent studies conducted at the village level using border discontinuity designs have yielded results which at times conflict with the existing literature, with landlord villages featuring *lower* poverty rates than non-landlord villages (Ratnoo, 2024); and British-ruled areas either better off (Colleoni, 2024) or no worse off (Lee, 2023). However, the methodological strengths of these studies must be weighed against their uncertain generalizability, as they are all limited to a small number of districts. Our study, in contrast, covers villages across 10 large states representing 70% of the country's rural population, allowing us to simultaneously avoid the pitfalls of ecological inference while reducing concerns of external validity within the context of India.

Another important factor driving inequality in India is the social structure, particularly inequities along caste cleavages. The most notable example of this is the socio-economic marginalization of the so-called "Schedules Castes" (SCs), which has deep roots in Indian history (Thangaraj, 1994). Among myriad other hardships, this group historically owned very little agricultural land, and was generally relegated to highly unfavorable tenancy arrangements and wage labor. Research has shown that, despite some improvements in recent years, this group continues to own little land (Tiwari et al., 2022). Another group facing significant marginalization is the "Scheduled Tribes" (STs) (Thangaraj, 1994), who often live in more rugged, forested areas, on the periphery of economically developed areas. While these groups are the most salient suffering such hardships, localized caste-baste inequities can be found across India (see, e.g., Anderson et al., 2015).

Since independence, a variety of land reforms have been introduced to ameliorate inequities arising from these historical factors. These have included laws eliminating the *zamindar* system, imposing ceilings on the amount of land owned, the reform or even abolition of tenancy arrangements, and consolidation of land holdings. While researchers have found some beneficial impacts of these reforms, the general consensus is that their impacts have been relatively modest, and in some cases actually harmful (Besley and Burgess, 2000; Besley et al., 2016).[12] In this paper, we do not study the impacts of land reforms, as they are implemented at the state level and are both numerous and diverse, and therefore not amenable to analysis in a sample of 10 states. Their impact is accounted for by the inclusion of state fixed effects; and the explanatory power of the latter in the Shapley decomposition exercise will in part capture the influence of land reforms.

[11] Two other works in this domain find better female outcomes in British-ruled districts compared to princely districts (Roy and Tam, 2025; Nandwani and Roychowdhury, 2023).

[12] Besley et al. (2016) note that tenancy reforms benefited richer and more productive middle-caste tenants but reduced land access for poorer low-caste tenants.

3 Data and Descriptive Statistics

3.1 Data Sources

We construct a comprehensive village-level dataset by integrating multiple data sources: the 2011 landholding data; demographic census data from 1991, 2001, and 2011; historical governance and land-tenure institutions; and spatial data capturing agricultural conditions, including water access, soil quality, and cropping patterns. Additionally, we incorporate market-access measures, such as proximity of villages to town, major roads, highways, railroads, local agricultural markets (*mandis*), and banks. We detail the data sources below.

3.1.1 Socio-Economic Caste Census 2011

The Socio-Economic Caste Census (SECC) was conducted in 2011, collecting information on household ownership of agricultural land in rural areas.[13] This is the first dataset providing land ownership measures at such a granular level, and to our knowledge is one of the richest data sets on land ownership for a developing country. The data is available for ten of the largest states of India—Punjab, Uttar Pradesh, Bihar, Rajasthan, Madhya Pradesh, Maharashtra, Karnataka, Tamil Nadu, Kerala, and West Bengal—accounting for approximately 75% of the rural population ($\sim$ 650 million people).[14] We compute village-level land distribution using the total agricultural land area owned.

3.1.2 Census: 1991, 2001, 2011

The last three demographic census rounds provide several village-level indicators that can be divided into two broad components. The first is related to population, including: total and male/female population, scheduled caste and scheduled tribe population, and the working population in the agriculture sector. The second is related public goods available in the village, including: roads, schools, health centers, and so on. The three rounds are combined at the village level, where 2001 and 2011 have better matching rates due to the cross-over files provided by the Census. The 1991 matching is based on name-based fuzzy matching. We also combined SECC 2011 with the Census 2011 dataset at the village level using name-based fuzzy matching.

[13]This excluded homestead land, which is land where the household lives, and which is not used for agricultural production.

[14]The dataset does not include the states of Andhra Pradesh, Telangana, Orissa, Chattisgarh, Jharkhand, Haryana, or Gujarat, nor some of the smaller states in the northeast and far north.

3.1.3 Pre-independence governance and land tenure systems

During the colonial period, different regions had different governance and land revenue collection systems in India. First, about one-third of the geographical area of India was not directly ruled by the British, but instead was included in the so-called "princely states," in which sovereignty was exercised by Indian princes with indirect control exercised by the British. In addition, within British-ruled districts, there were variations in the system of land revenue collections, which can be broadly classified into two—*zamindari* and *ryotwari*—with the main difference being that in *zamindari* areas, revenue collection happened through intermediaries and in *ryotwari* areas was directly collected from cultivators. These land-revenue systems in turn entailed important differences in land rights, with *zamindars* ultimately being granted ownership rights over large shares of the villages they administered, whereas cultivators retained ownership rights in *ryotwari* areas. In the princely states, rules related to land revenue collection were governed by Indian rulers and showed substantial variation.

The assignment to historical institutions come from Banerjee and Iyer (2005) and Iyer (2010), which assign princely state and land revenue status at the level of 1991 districts. As bookkeeping was less rigorous in princely states, information regarding land institutions remains unclear in these areas. We therefore combine all princely states into one to compare with directly ruled British areas; and divide British-ruled areas into the forementioned types of revenue collection zone.

In addition, in some analyses we make use of a new open source princely state map which seeks to precisely identify princely state boundaries, rather than making assignments based on 1991 district boundaries. To convert this map to a shapefile, we manually geo-referenced the coordinates choosing control points along the country's border and creating polygons tracing administrative boundaries of the princely states from the map. We then used this file to compute the distance of each Indian village from the nearest princely state boundary.[15]

3.1.4 Satellite-based and other spatial Data

We use a number of data sets derived from satellite imagery. These include: mean annual temperature and precipitation; altitude and ruggedness; soil suitability; and the presence of, or distance to, a water body. These data sets are taken from U.N. Food and Agricultural Organization's Global Agro Ecological Zones version 4 dataset (FAO's GAEZv4). Night light data comes from NOAA's

[15] The original map we used to create the shapefile is "1934-1947 — India Princely States Historical Maps 1934-47," extracted from an open source. 92% of villages in our sample have the same princely state status across the two data sources. Discrepancies between our map and that of Iyer (2010) come primarily from three sources: (1) errors in the open source map; (2) potential mismatches introduced by manual geo-referencing in our map; and (3) attribution by Iyer (2010) of princely status to entire 1991 districts where only a sub-region of the district had princely-state status.

National Geophysical Data Center's Defense Meteorological Satellite Program. We also use the MODIS Enhanced Vegetation Index (EVI), which captures dry season cultivation at a resolution of 1×1 sq kms (Jain et al., 2017).

Spatial data on the type of aquifer is taken from the India Water Resources Information System (WRIS). The Indian aquifer map includes 13 types of aquifer. We also include information on the location of command areas, which is taken from the India WRIS Wiki and the Central Water Commission's Management Information System of Water Resources Projects.

3.2 Descriptive Statistics: Land Inequality

Appendix Table A1 presents summary statistics for key variables. Column (1) gives the village mean. Columns (2)–(6) present the values at different percentiles (10, 25, 50, 75, and 90).

We focus on three key measures of inequality in this paper: (1) all-households land Gini; (2) landowner Gini (among landowning households); and (3) share of households who are landless. We additionally study other land distribution measures, including: share of land owned by the top 10%, top 5%, and top 1% of households; share of land owned by the wealthiest household in the village; and an indicator if a major landlord owns more than 50% of the agricultural land. We present statistics for several key land ownership variables disaggregated by state (Appendix Table A2).[16]

The mean village-level all-household Gini (on 0-100 scale) is 71.1, while the landowner Gini is a more modest 45.9. 46% of households are landless. The average size of a landholding is 6.2 hectares (excluding the landless population). However, 28.9% of land is held by households with 0–1 hectares, and 48.6% by households with 1–2 hectares. The mean share of land held by the largest landholder is 12.4%, while 3.8% of villages have a largest landholder who owns more than 50% of the land.

Appendix Figures A2a and A2b show the spatial distribution of inequality in India for all households and landed households, respectively. Appendix Figure A3 shows the distribution of Gini disaggregated by state. Appendix Table A2 provides the village-level averages for different land inequality measures we built. Looking at the the all-household Gini measure, Kerala exhibits the highest Gini coefficient (at 90), followed by Bihar, Punjab, Tamil Nadu and West Bengal, each with a Gini coefficient (around 80). Conversely, Karnataka and Rajasthan report the lowest Gini coefficient (below 65). Excluding landless population reduces the Gini coefficient for all states, and reduces variation across states, indicating that landlessness contributes significantly to the all-household Gini measure. [17]

[16] All the main results of the paper are also shown using top 10% land shares; these are reported in Appendix B at the end of the paper.

[17] A small note on the highest agricultural land inequality observed in Kerala is worth highlighting. First, our mea-

The mean agricultural landlessness within villages varies from 34% to 72% across Indian states. Among the four states where dependency on agriculture is particularly high, Rajasthan (34%) and Uttar Pradesh (39%) have a relatively lower level of landlessness than Madhya Pradesh (51%) and Bihar (59%). Punjab, known for its highly developed commercial agriculture sector, has the highest level of landlessness at 73%. The share of land owned by the top household ranges from a low of 7.3% in Uttar Pradesh, to a high of 20.1% in Bihar. Bihar and Kerala stand out for their high levels of land concentration, whether looking at the top 10%, 5%, or 1%; while Bihar and Punjab feature the highest share of villages in which a single landlord owns more than half of the available land (Column (8)).

4 Determinants of Land Inequality

In this section, we first look at the correlates of land inequality. Next, we perform a Shapley decomposition to understand the contribution of different factors to land inequality. We then turn to an analysis of individual elements in order to better understand the key drivers of land inequality. Following this, we conduct a heterogeneity analysis to understand how market access and structural transformation mediate the impacts of agricultural and historical factors on inequality.

4.1 Correlates of Land Inequality

We first perform a simple regression of land inequality on our variables of interest. For this exercise, we broadly classify our variables as agricultural, market and historical.[18] Agriculture-related variables ($\mathbf{A}_{vds}$) include: altitude; ruggedness; latitude; average annual number of days conducive to plant growth, precipitation, and temperature (all normalized); log distance from the nearest major river; an indicator of alluvial aquifer; a measure of soil suitability; and an indicator for being within a government irrigation scheme (or, "command area").[19] The market-related

sure is not due to data specificity. A recent paper using Agricultural Statistics from 2021, also finds the highest agricultural operational landholding inequality, Gini more than 90 (Kumar and Sharma (2024)). The reasons cited behind current high agricultural land inequality is multifaceted - large population has moved away from agriculture sector, change in cropping patterns with move towards coffee and rubber plantation agriculture, and post-independence land distribution was concentrated in homestead land. Lastly, the census definition of village is different - the average village size in Kerala is 17k, compared to 1.7k in Uttar Pradesh, which means more non-rural type areas are included in Kerala. Since Kerala has slightly more than 1k villages in the full sample of more than 250k villages, it doesn't change the main argument in subsequent analysis. Though not shown, all our results are robust to excluding Kerala.

[18] While this approach is motivated by Henderson et al. (2018), we make a number of modifications to make it more relevant in the context of India.

[19] Command areas are areas in which irrigation water is provided to villages via canals, either from upstream dams or river diversion. While command areas diverge from our other agricultural features in not being inherent attributes, they are included due to their significant impacts on agricultural productivity (Duflo and Pande, 2007; Blakeslee et al., 2023; Asher et al., 2024). In results not shown, we find that all results are robust to their exclusion.

variables ($\mathbf{M}_{vds}$) include: log distance to the nearest town; log distance to the nearest major road; and log distance to the nearest railroad. The historic variables ($\mathbf{H}_{vds}$) are: indicators for having been a princely state and for having the *zamindar* land tenure system in the past; as well as the percentage of the population that is Scheduled Caste (SC) and the percentage that is Scheduled Tribe (ST).

Table 1 gives the results. Among agricultural variables, we generally see that characteristics associated with greater agricultural productivity also have more inequality. However, given the substantial correlation between many of these variables, we avoid drawing strong conclusions about individual correlations. All market-access variables have negative coefficients, indicating that being farther from markets is associated with less inequality. Princely states are associated with lower land inequality, and *zamindar* areas with higher—though, due to the smaller within-state variation in these variables, they are sensitive to the inclusion of state fixed effects. Having a larger SC population is strongly associated with higher inequality, while there is no relationship with the size of the ST population once non-demographic covariates are included.

In Appendix Table A3 we perform the same exercise taking as the outome Gini amongst landowning households. The results of this analysis show similar patterns, though coefficients are generally smaller than before. One striking difference is that we now find that SC population has no impact on inequality, whereas ST population is associated with a large decline in inequality.

4.2 Shapley Decomposition

To understand the relative significance of these factors in determining land inequality, we next perform a Shapley decomposition using all-household Gini. The results are given in Table 2, Panel A.

We first perform this exercise excluding fixed effects (Columns (1)–(2)). This specification explains 18.3% of variation in inequality. Amongst the categories of explanatory variable, historic (39% of explained variation) is the strongest predictor of inequality, followed by agriculture-related (37%), and market-related (20%).

When we include state fixed effects (Columns (3)–(4)), the explanatory power of all three categories falls substantially, with state fixed effects accounting for 63% of explained variation. The decline in the explanatory power of historic variables is particularly pronounced, due to the fact that there is substantially less within-state variation in princely state or *zamindar* status. When we include district fixed effects (Columns (5)–(6)), the explanatory value of agriculture-related variables is close to zero, while districts fixed effects account for 85% of the explained variation, and the specification explains 34% of the total Gini variation.

In Panel B of Table 2, we perform the same exercise taking as the outcome the landowner Gini coefficient. Here we find that far less variation is explained overall, and less variation is explained by the different categories and by the different fixed effects. This finding is consistent with the decline in coefficient magnitude found in Appendix Table A3, and implies that much of the effect of these variables on inequality is due to impacts on landlessness.

4.3 Individual Components

In this section, we study the impacts of each of these factors using a variety of empirical approaches.

4.3.1 Agriculture

We first seek to understand the effect of agricultural productivity on inequality. For this exercise, we regress EVI—a measure of dry-season cropping using remotely sensed data—on the full vector of agriculture, historic, and market variables, and estimate the fitted values for the agriculture variables. We then convert this measure to the z-score of fitted values, and regress the Gini variable on this variable. Following equations describes it

$$EVI^{y}_{vds} = \alpha + \beta \mathbf{A}_{vds} + \gamma \mathbf{M}_{vds} + \theta \mathbf{H}_{vds} + \delta_s + \epsilon_{vds} \tag{1a}$$

$$Gini^{land}_{vds} = \alpha + \sum_{-2}^{1.5} \beta_i \widehat{\mathrm{EVI}}_{vds} + \theta ln(pop\ density)_{vds} + \gamma ln(total\ area)_{vds} + \delta_s + \epsilon_{vds}, \tag{1b}$$

where EVI^{y}_{vds} is the average of $y = 5$ years (2012–2016) of EVI to avoid any specific year's aberration, in village *v*, located in district *d* and state *s*. $\mathbf{A}_{vds}$, $\mathbf{M}_{vds}$, and $\mathbf{H}_{vds}$ are set of agricultural, market, and historic covariates, as listed in section 4.1. In equation 1b, $\widehat{\mathrm{EVI}}_{vds}$ is the normalized fitted agriculture value discretized at intervals of 0.25 from −2 to 1.5 sd, on which is regressed the all-household Gini, controlling for population density and total land area in logs in the village. State fixed effects (δ_s) are included, and error terms are clustered at the district level in both regressions.

Figure 2 plots the $\hat{\beta}_i$ coefficients and their associated standard errors. There is a sharp increase in inequality as land becomes more productive up to the 60th percentile (corresponding to a z-score of 0.25), after which inequality remains unchanged with further increases in agricultural productivity. This pattern is very similar across British-ruled and princely-state areas and is therefore not driven by historical institutional factors. To further understand this, we change the outcome variables in equation 1b to landless population share, and the share of land owned by different categories of landlords: marginal (0–1 ha), small (1–2 ha), semi-medium (2–4 ha), medium (4–10 ha), and large (>10 ha). The observed increase in the land inequality until the 60th percentile is predominantly driven by a combined effect of increasing landlessness, decreasing small (0–2 ha) and semi-medium (2–4 ha) landlords' land share, and a strong rise in the large (>10 ha) landlords' land share. Beyond the 60th percentile, the share of land owned by large landlords continues to increase, though with no impact on Gini, due to declining landlessness. Interestingly, there is almost no change in the medium-size (4–10 ha) land share throughout.

We complement this analysis by looking at the impact of (exogenously) being within an irrigation scheme, or "command area"—which is associated with increases in agricultural productivity (Blakeslee et al., 2023; Asher et al., 2024)—on inequality. Because the boundaries of command areas are primarily determined by topographical features and engineering considerations, whether a village is located within the irrigation zone is quasi-exogenous in close proximity to the boundary.[20] We therefore use a border RDD design to test whether being within a command area is associated with differences in inequality. The specification is given as

$$\begin{aligned} Gini^{land}_{vds} &= \alpha + \sum_{-10km}^{20km} \beta_i Dist\ Command\ Area_{vdsb} + \gamma \mathbf{A}_{vds} + \gamma ln(total\ area)_{vds} \\ &+ \delta_{ds} + \phi_b + \epsilon_{vds}, \end{aligned} \tag{2}$$

where $Gini^{land}$ is the all-household Gini, in village *v*, located in district *d* and state s. $\mathbf{A}_{vds}$is the set of agriculture-related covariates. $Dist\ Command\ Area_{vdsb}$ is the distance of the village boundary from the nearest command area boundary, which is negative if the village falls within the command area and positive if it is outside. We further control for total village land area in logs. δ_{ds} are district fixed effects, ϕ_b are the 10km boundary segment fixed effects, and standard errors are clustered at the command area project level.[21]

In Appendix Figure A4, we present the estimated $\hat{\beta}_i$ coefficients for the indicated distance intervals. Villages within the command area have higher levels of inequality, and a discontinuous jump at the border is evident. We present a consolidated impact in the Appendix Table A4, as well as the impact on agricultural productivity. There is an increase in the Gini coefficient of around 0.90 pp within the command area, which is significant at the 5% level, and is robust to varying bandwidths (10km, 20km) and the inclusion of the full set of historical covariates $\mathbf{H}_{vds}$.[22]

[20] Blakeslee et al. (2023) show that many boundaries, particularly those formed by canals, follow the base of hills, with sharp differences in geographic features on either side of the boundary; and therefore restrict the sample to boundaries in which differences are not present. We follow the same procedure here.

[21] Following Blakeslee et al. (2023), we further make sample restrictions: excluding boundary segments where the terrain gradient outside the command area is more than 1.5 degrees; removing villages which intersect the boundary; and removing villages for whom the nearest boundary is within 500 meters of a river.

[22] We exclude market-related variables ($\mathbf{M}_{vds}$) from this analysis, as command areas have been shown to influence the presence and scale of urban agglomerations (Blakeslee et al., 2023; Asher et al., 2024).

To the best of our knowledge, this is the first paper showing such a relationship between general land productivity (as opposed to suitability to *plantation* agriculture) and inequality. Two mechanisms within the existing literature may explain this result. First, following Braverman and Stiglitz (1989), if the attributes associated with great agricultural suitability are more readily exploited by the wealthy, then this would give rise to such a finding. Second, better agricultural suitability may increase the population, which will cause greater land fragmentation, and may do so differentially for the rich (Bardhan et al., 2014).[23] Two additional channels may also play a role. First, the higher population density of more agriculturally productive areas may bring about a more developed wage labor market, which would be of greater benefit to large landholders (Foster and Rosenzweig, 2022). In addition, agricultural productivity may have impacts on structural transformation (positive or negative), which in turn will affect the incentive to sell off land holdings, particularly for smallholders with non-viable plots.

4.3.2 History: Princely States and *Zamindar*

In this section, we seek to better understand the drivers of the relationship between institutions and inequality. We additionally assess the causal validity of this relationship using a spatial regression discontinuity design (RDD).

We first investigate the impact of institutions on different measures of inequality, starting with a simple regression:

$$Land\ Ineq^{i}_{vds} = \alpha + \Phi^{i} Institution_{vds} + \beta \mathbf{A}_{vds} + \gamma \mathbf{M}_{vds} + \theta \mathbf{H}_{vds} + \delta_s + \epsilon_{vds}, \tag{3}$$

where $Land\ Ineq_{vds}$ is the land inequality measure (i) in village *v*, district *d*, and state *s*. $Institution_{vds}$ is the main treatment variable, equaling 1 for villages within the princely state (*zamindar* area) and 0 otherwise. The princely state variable comes from Iyer (2010), which defines princely state status at the district level using 1991 district boundaries; while the *zamindar* variables comes from Banerjee and Iyer (2005), again using 1991 districts. Because we use 2011 districts, institutional variables can (rarely) vary within districts. $\mathbf{A}_{vds}$, $\mathbf{M}_{vds}$, and $\mathbf{H}_{vds}$ are sets of agricultural, market, and historic covariates (other than the main treatment variable), as listed in section 4.1. Several different land inequality measures are taken as the outcome : all-household Gini; landowner Gini; percent of landless household; percent of household owning; percent of land owned by those with <2 hectares of land, 2-4 hectares, 4-10 hectares, and >10 hectares land; percent of land owned by the wealthiest household; and indicators (=**100**) for villages where the wealthiest household owns more than 30% or 50% of total agricultural village land, respec-

[23] This may happen if the wealthy have fewer children; or if the division of land amongst sons causes landholdings of the poor to reach a non-viable size faster, and so be sold off to wealthier households.

tively.

We run separate regressions for each land inequality measure, and Table 3 presents the coefficient estimates ($\hat{\Phi}^i$) along with standard errors from different regressions. We present this analysis with and without state fixed effects (δ_s), because the historical institutional variables are highly correlated with state fixed effects. We also conduct an analysis excluding state fixed effects while including a vector of controls to capture disparities in development plausibly attributable to inter-state policy and institutional differences.[24]

Princely states exhibit lower levels of land inequality amongst all households. This is driven primarily by a lower share of landless households, as there is no significant impact when examining the landowner Gini. When disaggregated by farm size, the findings indicate fewer landlords holding moderate-sized farms (4-10 hectares), coupled with a slight increase in very large farms ($>$10 hectares). Despite this increase in larger farms, the net effect on overall land inequality (as measured by the Gini coefficient) reflects that these areas have lower inequality. Furthermore, the increased likelihood of a dominant landlord presence, indicated by the last three outcomes, suggests that reduced inequality in princely states is not merely due to the absence of large landholders. In contrast, *zamindar* areas clearly show higher levels of land inequality, driven by a marked reduction in the share of small farmholders and a notable increase in the presence of dominant landlords.

The greater inequality in *zamindar* areas is driven by both inequality among landowners and the percentage of landlessness. Without state fixed effects, we see an increase in both Gini measures– all-households and within landowners– though the latter is insignificant with the inclusion of state fixed effects. There is a decline in the share of land held by those with less than 10 hectares, and a large increase in the share of land held by those with more than 10 hectares. Approximately two-thirds of the increase in large holdings is driven by the share of land held by the wealthiest household. This is associated with a greater incidence of landlord-dominated villages in *zamindar* areas.

The results are sensitive to the inclusion of state fixed effects, which is due to the lack of within-state variation in institutional status, particularly for *zamindar* tenure.[25] When excluding fixed effects and including controls for broader measures of development (Columns 3 and 6), we find the estimates to be remarkably robust, increasing our confidence in the causal impact of these institutional features.

[24]These are: ln(light density); the share of workers outside of agriculture; the presence of a paved road; ln(number of workers) in firms with $10 - 50$ workers; and ln (number of workers) in firms with > 50 workers. In results not shown, we also include ln(population density), which yields virtually identical results.

[25]Banerjee and Iyer (2005) excluded state fixed effects for this very reason. The authors performed robustness tests that included state fixed effects.

Border Discontinuity Design To more rigorously establish the impact of institutions, we next turn to a border discontinuity design. For this analysis, we compare all-household Gini for places in the close proximity (10 and 20 km) on either side of the relevant institutional feature (British-vs-non-British, *zamindar*-vs-non-*zamindar*). We use the following spatial regression discontinuity (RD) design:

$$Gini^{land}_{vdsb} = \alpha + \Phi Institution_{vds} + \beta \mathbf{Z}_{vds} + \delta_s + \Delta_b + \epsilon_{vds}, \tag{4}$$

where $Gini^{land}_{vdsb}$ is the land Gini in a given village *v*, located in district *d* and state *s*. $Institution_{vds}$ is same as defined earlier. γ_b are boundary segment fixed effects for each 20 km interval along the boundary separating villages inside and outside the areas possessing the respective institutional features. State fixed effects δ_s are included in the princely state (but not *zamindar*) regressions to account for state-level shocks, including post-independence land reforms, which varied by state (Besley and Burgess, 2000), and standard errors are clustered at the district level. We also gradually control for other determinants of land inequality $\mathbf{Z}_{vds}$, including the full set of agriculture and historic (caste) variables.

Table 4 gives the results for princely states, using a variety of bandwidths, fixed effects, and controls. For this analysis, we define princely state status and boundary distance using our newly-created maps, but dropping any villages where princely state status differs from that of Iyer (2010).[26] Results are similar using 20km or 10km bandwidths, showing an approximately 2.7 pp lower Gini within princely states (Columns 1 and 2), though 10km bandwidth yields slightly smaller coefficients. The inclusion of full set of controls successively—agricultural ($\mathbf{A}_{vds}$) in Panel B, the other historic variables (i.e., caste shares) in Panel C, and market-related variables ($\mathbf{M}_{vds}$) in Panel D—reduce the coefficient to 1.9 pp. In Columns (3) and (4), we add the princely state fixed effects, which removes the time invariant characteristics in the land inequality around specific princely state boundaries; and Columns (5) and (6) include boundary segment fixed effects. The latter two approaches show an approximately 2.0 pp decline in inequality in princely states, significant at the 1% level. The coefficients are relatively stable across specifications, becoming only slightly smaller as more controls are included.

Table 5 presents the results using the explanatory variable *zamindar* land tenure system. For this analysis, we exclude princely states from the sample. Due to sample limitations,[27] we focus on the UP-Bihar boundary. Because treatment status is collinear with state fixed effects, we exclude the latter and instead include the aforementioned supplemental controls to account for state-

[26]When restricting the sample to within 20kms of the princely state boundary using our maps, 82% of villages have the same princely state status using our maps or those of Iyer (2010). Our results are robust to alternative definitions of princely-state status and distance using only our maps or only those of Iyer (2010).

[27]We lack inequality data along large segments of the *zamindar* boundary, and other portions of the boundary coincide with princely state boundaries.

level differences in policies and institutions. We see that the Gini coefficient in *zamindar* areas is approximately 3–5 percentage points higher; and that again this result is remarkably robust across specifications.

To further bolster the validity of the research design, we present figures showing the relationship between inequality and distance to the respective institutional boundaries. This follows the specification used in the command area analysis (equation 2). Figure 3 presents the coefficients for the indicated distance intervals from the princely state and *zamindar* boundaries, respectively. The impacts are large and discontinuous at the respective boundaries, where we see an approximately 2 pp decline in all-household Gini in princely states, and a 5 pp increase in *zamindar* areas.

4.3.3 History: Caste

Caste plays an important role in the structure of land ownership in India. Most land has historically been owned by the upper castes of society, with the lower castes occupying the status of landless laborers and tenant farmers (Mohanty, 2001). Land reforms after independence attempted to redress some of these historical inequities, but were generally deemed to have brought about little improvement (Besley and Burgess, 2000; Besley et al., 2016). While recent years have seen some increase in land holdings among lower ranks, historical inequities continue to persist (Deshpande, 2001; Bharti, 2018; Bharti et al., 2024).

We begin with the following to understand relationship between the SC population and land inequality

$$\begin{aligned} Land\ Ineq_{vds} &= \alpha + \sum_i \beta_i \mathbf{1}[100 * SC\ pop\ share_{vds} \in \mathcal{B}_i] + \gamma \mathbf{Z}_{vds} + \theta ln(total\ area)_{vds} \\ &+ \delta_{ds} + \epsilon_{vds}, \end{aligned} \tag{5}$$

where $Land\ Ineq_{vds}$ is the inequality measure (all-household Gini, landowner Gini and share of landlessness) in village *v*, located in district *d* and state *s*. The percentage of the SC population is discretised at intervals of 5 pp such that $\mathcal{B}_i = \{(5-10]\%, (10-15]\%, ..., (95-100]\%\}$. We further control for total village land area in logs. $\mathbf{Z}_{vds}$ is the full set of agriculture, market, and historic covariates (excluding SC population share). δ_{ds} are district fixed effects, and standard errors are clustered at the district level.

Figure 4 depicts the coefficient estimates ($\hat{\beta}_i$) and confidence intervals for all-household Gini. The relationship for all-household Gini is positive and largely linear in the share of the population being SC between 5–80%; is far steeper when the share increases from 5–10%; and is negative for values greater than 80%. This relationship is entirely driven by landlessness, with the relationship between the landowner Gini and the SC population in fact being somewhat negative.

In Panel A of Appendix Table A5, we estimate the linear relationship between all-household Gini and SC population share ($SC\ pop\ share_{vds}$) under a variety of specifications. Column (1) is a simple correlation; in Column (2), we include the full set of agriculture, history, and market controls (i.e. $\mathbf{Z}_{vds}$); and in Columns (3)–(5) add state, district, and subdistrict fixed effects, respectively. Standard errors are clustered at the district level. The coefficient is positive (12.2-12.9), statistically significant (at 1% level), and is remarkably stable across specifications. The effect size indicates that moving from a village at the 25^{th} percentile of the SC population ($SC\ pop\ share_{vds} = 0.012$) to the 75^{th} percentile ($SC\ pop\ share_{vds} = 0.272$) increases the all-household Gini by 4.4% of the mean (71.2).

Because migrant agricultural workers are often from the SC population, one might be concerned that the landless population, and hence the all-household Gini, may be driven by migration patterns. In Appendix Table A6, we regress population and labor force composition on the SC population share. We find that population density is indeed higher when district fixed effects are included, and that this is associated with a larger agricultural wage-labor population. Hence, in Panel B of Appendix Table A5, we estimate SC regressions controlling for the (ln) population density, which yields virtually identical results.

We test for heterogeneity in this relationship across states. To compare the estimates across states, we standardized the SC population share variable by the respective states' means and standard deviations. The coefficients for regressions run separately for individual states are plotted in Appendix Figure A8. For most states, increasing the SC population by 1 standard deviation increases the Gini between 3–5 pp and is statistically significant at 1% level. In Punjab and Maharashtra, the states in which agricultural production is more market-oriented, the coefficients are the highest. One notable exception to this pattern are the states of Kerala and West Bengal, which were long governed by left-wing parties, and in which post-independence land reforms are generally considered to have been the most successful: in these states, the relationship between SC population share and land inequality is substantially smaller and is statistically insignificant.

4.3.4 Market Integration

We next examine the relationship between the market (access)-related variables and land inequality. For this, we use the specification:

$$Gini^{land}_{vds} = \alpha + \sum_i \beta_i \mathbf{1}[Dist\ Market_{vds} \in \mathcal{B}_i] + \gamma \mathbf{Z}_{vds} + \theta ln(total\ area)_{vds} + \delta_{ds} + \epsilon_{vds}, \quad (6)$$

where $Gini^{land}_{vds}$ is the all-household Gini in village *v*, district *d*, and state *s*. Dist Market is the distance of the village from the respective market features, discretized at intervals of 2.5 km such that $\mathcal{B}_i = \{(0 - 2.5]km, (2.5 - 5]km, ..., (32.5 - 35]km\}$, resulting in fourteen dummies, and generating thirteen $\hat{\beta}_i$'s. Since we have three market-related variables—distance to major roads, towns and railroads—we run separate regressions for each. The rest of the specification is similar to before. Figure 5 plots the coefficient estimates.

For each market-related variable, the Gini coefficient increases substantially as proximity to the respective locations decreases. The increase is the largest for towns; for roads and railway stations the relationship is smaller, but still positive. Moreover, we see that inequality is elevated up to further distances from towns (10kms) than distances to roads and railway stations (2.5kms). In Appendix Figure A5, we additionally control for population density, allowing us to isolate the impacts of market integration from the impacts coming from population density. This specification yields the same patterns, though impacts are somewhat smaller.

In Appendix Figure A6 we look separately at national highways which were significantly modernized after 2000: the Golden Quadrilateral, which had been largely completed by 2006, five years prior to the land survey; and North-South-East-West (NSEW) expansion, of which only 10% had been completed by 2006, and 81% completed at the time of the land survey (Ghani et al., 2016). In addition, the Golden Quadrilateral was generally situated along long-established major highways, whereas the NSEW was established along historically less important routes. For the Golden Quadrilateral, there is an increase in land inequality up to 25 kms of the highway, which increases by 2 pp in the immediate vicinity of the highway. In contrast, there is no evidence for greater inequality beyond 5 kms of the NSEW highway, and even within 5 kms of the highway, the impact is approximately 0.75 pp. These results are consistent with those of Berg et al. (2023), who test the impacts of these road networks on district-level inequality, finding that a 10% increase in market access increases the Gini by 2.55% and the landlessness Gini by 6.78%.

We provide further evidence for the role of markets by studying the relationship between inequality and two other types of market integration: the presence of banks and the presence of government-sanctioned agricultural markets (*mandis*). Because these sites may be located in places where there is more commercial activity, which may itself affect land inequality, we include supplemental controls, consisting of: (ln) population density, (ln) light density, the share of the non-agricultural workforce, and employment in medium and large firms. Appendix Figure A7 shows the relationship between distance to villages possessing these facilities and all-household Gini. We see that both bank presence and agricultural markets are associated with greater land inequality, with little evidence for spillovers to villages even very close to the facilities.

4.4 Impact of Structural Transformation on Agricultural and Historical Factors

We next seek to understand how economic modernization is likely to influence the role of agriculture and history in shaping inequality. There are intuitive reasons to think market access should influence the role of first-nature factors on land inequality, as the returns to land ownership come to align with the economic influence of non-agricultural production. A similar logic may predict that these market forces would also attenuate the role of history, though the latter may generate forms of inequality that are more resistant to change. We therefore explore whether proximity to such employment opportunities mediates the impact of agricultural productivity and historical factors on patterns of land inequality.

To examine this, we first interact an indicator for structural transformation with the fitted agriculture productivity ($\widehat{\text{EVI}}_{vds}$) and our historic variables ($\mathbf{H}_{vds}$) in Table 6. We define structural transformation using an indicator for sub-districts in which the share of the working population outside of agriculture is more than one standard deviation above the mean. In Columns (1)–(2), we include state fixed effects, and in Columns (3)–(4), district fixed effects. To ensure we are not simply capturing the effect of proximity to a town, we include a town-proximity indicator, along with its interaction with the fitted agriculture and historical variables. We additionally include a control for (ln) population density in Column (4).

We find that places with a large non-agricultural workforce are substantially more unequal (3–7 percentage points higher Gini). In addition, we find that the influence of agricultural potential on inequality is reduced by 50–100% in these places (interaction coefficient is negative and significant). In contrast, there is no decline in the role of historic factors in high-manufacturing areas (all interaction coefficients are insignificant).

In Appendix Table A7, we test for the impact of town proximity on agricultural and historic factors. Drawing on the insights of Appendix Figure A5, we define the town proximity variable to be an indicator equal to 1 if a village is within 10kms of any town. We find that the direct impact of fitted agriculture ($\widehat{\text{EVI}}_{vds}$) is mitigated by approximately 1/3 in close proximity to a town. As before, all interaction coefficients with historical factors are insignificant, highlighting the persistence of historical factors in determining inequality.

Appendix Figure A9 provides further insights into the relationship between town proximity and the influence of agriculture on inequality. The coefficients are from the interaction terms of (fitted) agriculture and indicators for the respective distance bins from the nearest town. This graph shows the decline in the relationship between agriculture and inequality in close proximity to towns, as well as a sharp increase at distances greater than 30kms.

We next seek to understand the influence of all structural transformation and market-related variables on the influence of agricultural factors. (We omit an analysis of their influence on historic variables, given the lack of interaction effects in the previous analysis.) Appendix Table A8 present the results. We see that the non-agricultural labor force and town proximity have independent impacts on the influence of agricultural productivity; and that the presence of a *mandi* also reduces the influence of agricultural productivity. However, we find no evidence that proximity to a road or railway station, or the presence of a bank, affects the influence of agricultural productivity.

These findings indicate that the impacts of agricultural suitability on inequality predominate only where market access is weak and structural transformation is less advanced. As such, it poses the key question of whether the findings for India would apply in the context of more advanced, spatially integrated economies with a smaller agricultural sector. The lack of an association between historical variables and structural transformation suggests that the historical drivers of inequality are less susceptible to mitigation by market operations.

5 Inequality and Public Goods

An extensive literature has argued that inequality drives poor provision of public goods, particularly education, and that this ultimately undermines broader economic development (Engerman and Sokoloff, 2000; Galor et al., 2009; Goñi, 2023). A related literature on clientelism and public goods similarly finds that where inequalities are stark, local elites may wield disproportionate political influence, which is associated with worse public goods provision (Acemoglu et al., 2014; Anderson et al., 2015). However, in some contexts inequality has been found to be associated with higher levels of public goods provision, which is attributed to the presence of a local elite capable of solving collective action problems or lobbying for government resources (Dell, 2010; Banerjee and Somanathan, 2007). Banerjee and Somanathan (2007) show that this is true specifically of India, where high land (possession) concentration in electoral districts is associated with higher levels of public goods provision.

In this section, we test for the relationship between land inequality (Gini and landlord dominance) and public goods provision.[28] Because public goods may be endogenous to land inequality, we focus on public goods that are *prima facie* unlikely to have contributed significantly to inequality, and control for broader levels of development that may drive both public goods and inequality.

In Appendix Figure A10, we plot the relationship between four types of public goods—

[28] While some researchers have sought to estimate the relationship between inequality and economic development, this is infeasible in our case, as the drivers of inequality also have direct impacts on economic development, and our inequality data consists of a single cross section from a relatively late date (2011).

presence of a government primary school, paved road, primary health clinic, and sanitation campaign—and all-household Gini. While the raw correlation is generally upwards sloping and concave, the inclusion of controls leads to an inverted-U relationship. This suggests that intermediate levels of inequality may promote the provision of public goods, but that extreme inequality is less favorable, and sometimes even detrimental, to public goods provision.

We run the following basic model motivated by the literature on public goods (specifically, Banerjee and Somanathan, 2007):

$$\begin{aligned} Public\ Good^i_{vds} &= \alpha + \beta^i_1 Gini_{vds} + \beta^i_2 Gini^2_{vds} + \gamma \mathbf{Z}_{vds} + \sigma^i * \mathbf{s}_{vds} + \theta ln(pop\ density)_{vds} \\ &+ \delta_{ds} + \epsilon_{vds}, \end{aligned} \tag{7}$$

where $Public\ Good^i_{vds}$ is the availability of ith public good in village *v*, in district *d*, and state *s*. $Gini_{vds}$ is the land inequality measure. When using landowner Gini, we additionally control for landlessness population share. $Gini^2_{vds}$ is the quadratic term due to the observed inverted-U relationship. $\mathbf{Z}_{vds}$ is the full set of agriculture, market and historic-related covariates. $\mathbf{s}_{vds}$ is index of social fractionalization, computed as $1 - SC\ pop\ share^2 - ST\ pop\ share^2 - Oth\ pop\ share^2$, the inclusion of which follows Banerjee and Somanathan (2007). δ_{ds} is fixed effect absorbing the time-invariant district characteristics, including historical land revenue. Finally, we cluster the standard errors at the district level.

In Appendix Table A9, we present these results. In Columns (1) and (2), Gini is measured using all households; and in Columns (3)-(6) using only landowning households. We find that public goods are positively associated with the land inequality, similar to the findings of other researchers at higher administrative levels (Banerjee and Somanathan, 2007). However, the relationship is non-linear, with negative and significant $\hat{\beta}^i_2$ coefficients for the quadratic term. We furthermore find that public goods are positively associated with the share of landless households, suggesting that the government is targeting the most deprived populations.

In Table 7, we test for the impact of having a landlord, which is defined as an individual owning more than 30% of the land. We repeat the analysis from before, controlling for Gini flexibly. There is a decline in all public goods when the village is dominated by a landlord. The decline is largest for government primary schools, which are 10 pp less likely to be present in a landlord-dominated village. These results are robust to the inclusion of both a linear and cubic (Columns (2) and (4)) in Gini, and controlling for the share of population in the non-agricultural sector (Columns (5)- (6)).

The finding that inequality is positively associated with public goods provision appears to contradict existing research (Engerman and Sokoloff, 2000; Easterly, 2007), but is consistent with the findings of Banerjee and Somanathan (2007) on India. As the latter argue, this may indicate that inequality is associated with the presence of a local elite that is better able to lobby for government resources and benefits from public goods. However, when there is a single dominant landlord, this mechanism breaks down, suggesting that an individual landlord may have less interest in the village's welfare, perhaps because such landlords are often absentee landlords (Lee, 2023).

6 Conclusion

This paper provides some of the first evidence for the drivers on land inequality in developing countries. We find that agricultural suitability, market access, and history all have important impacts on inequality, and collectively account for 18.3% of variation in land inequality. Agricultural suitability and history are of approximately equal importance, explaining 38% and 43% of inequality patterns, respectively, followed by markets (19%).

Agricultural productivity has a concave relationship with inequality, increasing montonically until the 60th percentile of agricultural productivity, then showing no relationship at higher levels. This is driven by the expansion of large farms at the expense of small farms, and a rise in landlessness. Amongst institutions, the *zamindari* system, which redistributed ownership rights under the British to a small group of landlords, is strongly associated with modern land ownership inequality; which is driven by substantial increases in the share of land held by the wealthiest household, reductions in shares held by smallholders, and an increase in landlessness. Rule by the British, as opposed to indigenous royal households, is also associated with an increase in present-day inequality, though the impact is far more modest. The presence of a large SC population is also strongly associated with greater inequality, primarily driven by heightened landlessness.

We furthermore find that structural transformation largely subsumes the influence of agricultural suitability on inequality, while market access reduces the effect of agriculture by 1/3. However, structural transformation and market integration have no impact on the historic drivers of inequality, suggesting these may be more resilient to the influence of the modern economy.

Finally, we also find evidence for a positive, concave relationship between inequality and public goods availability, with some public goods becoming slightly negative at high levels of inequality. This suggests that some degree of inequality is useful for securing public goods, whether because the more affluent class is better able to solve collective action problems or because it is better able to lobby the government for resources. However, where a dominant landlord is present, there is a sharp decline in public goods provision, particularly education.

References

Acemoglu, D., T. Reed, and J. A. Robinson (2014): "Chiefs: Economic development and elite control of civil society in Sierra Leone," *Journal of Political Economy*, 122, 319–368.

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

——— (2020): "Land reform and productivity: A quantitative analysis with micro data," *American Economic Journal: Macroeconomics*, 12, 1–39.

Alfani, G. (2010): "The effects of plague on the distribution of property: Ivrea, Northern Italy 1630," *Population Studies*, 64, 61–75.

——— (2022): "Epidemics, inequality, and poverty in preindustrial and early industrial times," *Journal of Economic Literature*, 60, 3–40.

Allen, T. and D. Atkin (2022): "Volatility and the Gains from Trade," *Econometrica*, 90, 2053–2092.

Anderson, S., P. Francois, and A. Kotwal (2015): "Clientelism in Indian villages," *American Economic Review*, 105, 1780–1816.

Asher, S., A. Campion, D. Gollin, and P. Novosad (2024): "The Long-run Development Impacts of Agricultural Productivity Gains: Evidence From Irrigation Canals in India." *Working Paper*.

Assunção, J. (2008): *9 Land Reform and Land Holdings in Brazil*, Oxford University Press.

Asturias, J., M. García-Santana, and R. Ramos (2019): "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 17, 1881–1940.

Banerjee, A. and L. Iyer (2005): "History, institutions, and economic performance: The legacy of colonial land tenure systems in India," *American economic review*, 95, 1190–1213.

Banerjee, A. and R. Somanathan (2007): "The political economy of public goods: Some evidence from India," *Journal of development Economics*, 82, 287–314.

Bardhan, P., M. Luca, D. Mookherjee, and F. Pino (2014): "Evolution of land distribution in West Bengal 1967–2004: Role of land reform and demographic changes," *Journal of Development Economics*, 110, 171–190.

Bauluz, L., Y. Govind, F. Novokmet, and D. S. Ordonez (2025): "Land Inequality in The Developing World," *Working Paper*.

Bequet, L. (2024): "Agricultural Productivity and Land Inequality: Evidence from the Philippines," *The World Bank Economic Review*, 38, 229–250.

Berg, C. N., B. Blankespoor, M. S. Emran, and F. J. Shilpi (2023): "Does Market Integration Increase Rural Land Inequality? Evidence from India," *World Bank Policy Research Working Paper 10286.*

Besley, T. and R. Burgess (2000): "Land reform, poverty reduction, and growth: Evidence from India," *The Quarterly Journal of Economics*, 115, 389–430.

Besley, T., J. Leight, R. Pande, and V. Rao (2016): "Long-run impacts of land regulation: Evidence from tenancy reform in India," *Journal of Development Economics*, 118, 72–87.

Bharti, N. K. (2018): "Wealth inequality, class and caste in India, 1961-2012," *WID.world WORKING PAPER.*

Bharti, N. K., L. Chancel, T. Piketty, and A. Somanchi (2024): "Towards Tax Justice and Wealth Redistribution in India," *Economic & Political Weekly*, 59, 91.

Binswanger-Mkhize, H. P., C. Bourguignon, and R. J. E. van den Brink (2009): *Agricultural land redistribution: toward greater consensus*, World Bank Publications.

Blakeslee, D., A. Dar, R. Fishman, S. Malik, H. S. Pellegrina, and K. S. Bagavathinathan (2023): "Irrigation and the spatial pattern of local economic development in India," *Journal of Development Economics*, 161, 102997.

Boberg-Fazlić, N., M. Lampe, P. M. Lasheras, and P. Sharp (2022): "Winners and losers from agrarian reform: Evidence from Danish land inequality 1682–1895," *Journal of Development Economics*, 155, 102813.

Bogaard, A., A. Styring, J. Whitlam, M. Fochesato, and S. Bowles (2017): "The agroecology of (in) equality: Integrating Gini coefficients with recent bioarchaeological work on early urbanization in northern Mesopotamia and southwest Germany," *Quantifying Ancient Inequality: The Archaeology of Wealth Differences.*

Bowles, S. and M. Fochesato (2024): "The origins of enduring economic inequality," *Journal of Economic Literature*, 62, 1475–1537.

Braverman, A. and J. E. Stiglitz (1989): *Credit rationing, tenancy, productivity, and the dynamics of inequality*, vol. 176, World Bank Publications.

Caum-Julio, J. (2024): "Can colonial institutions explain differences in labour returns? Evidence from rural colonial India," *The Economic History Review*, 77, 288–316.

Chaudhury, P. D. (2005): "Modal split between rail and road modes of transport in India," *Vikalpa*, 30, 17–34.

Colleoni, M. (2024): "The long-term welfare effects of colonial institutions: Evidence from Central India," *Journal of Development Economics*, 166, 103170.

De Janvry, A., K. Emerick, M. Gonzalez-Navarro, and E. Sadoulet (2015): "Delinking land rights from land use: Certification and migration in Mexico," *American Economic Review*, 105, 3125–3149.

de Janvry, A. and E. Sadoulet (1989): "A study in resistance to institutional change: The lost game of Latin American land reform," *World Development*, 17, 1397–1407.

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

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

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

Deshpande, A. (2001): "Caste at birth? Redefining disparity in India," *Review of Development Economics*, 5, 130–144.

Donaldson, D. (2018): "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American economic review*, 108, 899–934.

Donaldson, D. and R. Hornbeck (2016): "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 131, 799–858.

Duflo, E. and R. Pande (2007): "Dams," *The Quarterly Journal of Economics*, 122, 601–646.

Easterly, W. (2001): "The middle class consensus and economic development," *Journal of economic growth*, 6, 317–335.

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

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

Erickson, L. A. (2004): "Dimensions of land inequality and economic development," *IMF Working Paper*.

Fafchamps, M. and F. Shilpi (2005): "Cities and specialisation: evidence from South Asia," *The Economic Journal*, 115, 477–504.

FAO (2014): "The State of Food and Agriculture," *Food and Agriculture Organization of the United Nations*.

Fenske, J., N. Kala, and J. Wei (2023): "Railways and cities in India," *Journal of Development Economics*, 161, 103038.

Finley, T., R. Franck, and N. D. Johnson (2021): "The effects of land redistribution: Evidence from the French Revolution," *The Journal of Law and Economics*, 64, 233–267.

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, 636–680.

Frankema, E. (2005): *The Colonial Origins of Inequality: Causes and Consequences of Land Distribution*, Iberoamerica Institut fur Wirtschaftforchung der Universitat Gottingen.

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

Galli, S. and K. Rönnbäck (2021): "Land distribution and inequality in a black settler colony: The case of Sierra Leone, 1792–1831," *The Economic History Review*, 74, 115–137.

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, 143–179.

Ghani, E., A. G. Goswami, and W. R. Kerr (2016): "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing," *The Economic Journal*, 126, 317–357.

Goñi, M. (2023): "Landed elites and education provision in England: evidence from school boards, 1871-99," *Journal of Economic Growth*, 28, 125–171.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018): "The global distribution of economic activity: nature, history, and the role of trade," *The Quarterly Journal of Economics*, 133, 357–406.

Hornbeck, R. and M. Rotemberg (2024): "Growth off the rails: Aggregate productivity growth in distorted economies," *Journal of Political Economy*, 132, 3547–3602.

İşcan, T. B. (2018): "Redistributive land reform and structural change in Japan, South Korea, and Taiwan," *American Journal of Agricultural Economics*, 100, 732–761.

Iyer, L. (2010): "Direct versus Indirect Colonial Rule in India: Long-Term Consequences," *The Review of Economics and Statistics*, 92, 693–713, publisher: MIT Press.

Iyer, L. and C. Weir (2025): "The Colonial Legacy in India: How Persistent Are the Effects of Historical Institutions?" *Journal of Development Economics*, 103576.

Jain, M., P. Mondal, G. L. Galford, G. Fiske, and R. S. DeFries (2017): "An automated approach to map winter cropped area of smallholder farms across large scales using MODIS imagery," *Remote Sensing*, 9, 566.

Jayne, T. S., J. Chamberlin, L. Traub, N. Sitko, M. Muyanga, F. K. Yeboah, W. Anseeuw, A. Chapoto, A. Wineman, C. Nkonde, et al. (2016): "Africa's changing farm size distribution patterns: the rise of medium-scale farms," *Agricultural Economics*, 47, 197–214.

Jha, P. and K. Talathi (2024): "Impact of Colonial Institutions on Economic Growth and Development in India: Evidence from Night-Lights Data," *Economic Development and Cultural Change*, 72, 000–000.

Kumar, A. and T. Sharma (2024): "Extent of Inequality in the Distribution of Land Holdings in India: A State-Wise Analysis," *International Journal of Multidisciplinary Research and Publications*, 6, 88–108.

Lee, A. (2023): "Historical inequality at the grassroots: Local public goods in an Indian district, 1905–2011," *Comparative Political Studies*, 56, 1824–1857.

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.

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

Markevich, A. and E. Zhuravskaya (2018): "The economic effects of the abolition of serfdom: Evidence from the Russian Empire," *American Economic Review*, 108, 1074–1117.

Misra, K. (2019): "Does historical land inequality attenuate the positive impact of India's employment guarantee program?" *World Development*, 124, 104662.

Mohanty, B. B. (2001): "Land distribution among scheduled castes and tribes," *Economic and Political Weekly*, 3857–3868.

Moshrif, R. (2025): "Long-term Land Inequality and Post-Colonial Land Reform in Egypt (1896-2020)," Tech. rep., HAL.

Nandwani, B. and P. Roychowdhury (2023): "British colonialism and women empowerment in India," Tech. rep., GLO Discussion Paper.

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

Nunn, N. (2020): "The historical roots of economic development," *Science*, 367, eaaz9986.

Rao, M., J. Eberhard, and P. Bharadwaj (2022): "Towns and Rural Land Inequality in India," Tech. rep., Working Paper.

Ratnoo, V. D. (2024): "Persistent effects of colonial land tenure institutions: Village-level evidence from India," *Journal of Development Economics*, 168, 103247.

Roy, S. and E. H. Tam (2025): "Impact of British colonial gender reform on early female marriages and gender gap in education: Evidence from child marriage abolition act, 1929," *World Inequality Lab Working Paper Series.*

Roy, T. (2014): "Geography or politics? Regional inequality in colonial India," *European Review of Economic History*, 18, 324–348.

Sellars, E. A. and J. Alix-Garcia (2018): "Labor scarcity, land tenure, and historical legacy: Evidence from Mexico," *Journal of Development Economics*, 135, 504–516.

Thangaraj, M. (1994): "Access to Land for the Scheduled Castes and Scheduled Tribes in India," *Madras Institute of Development Studies, Working Paper No. 122*.

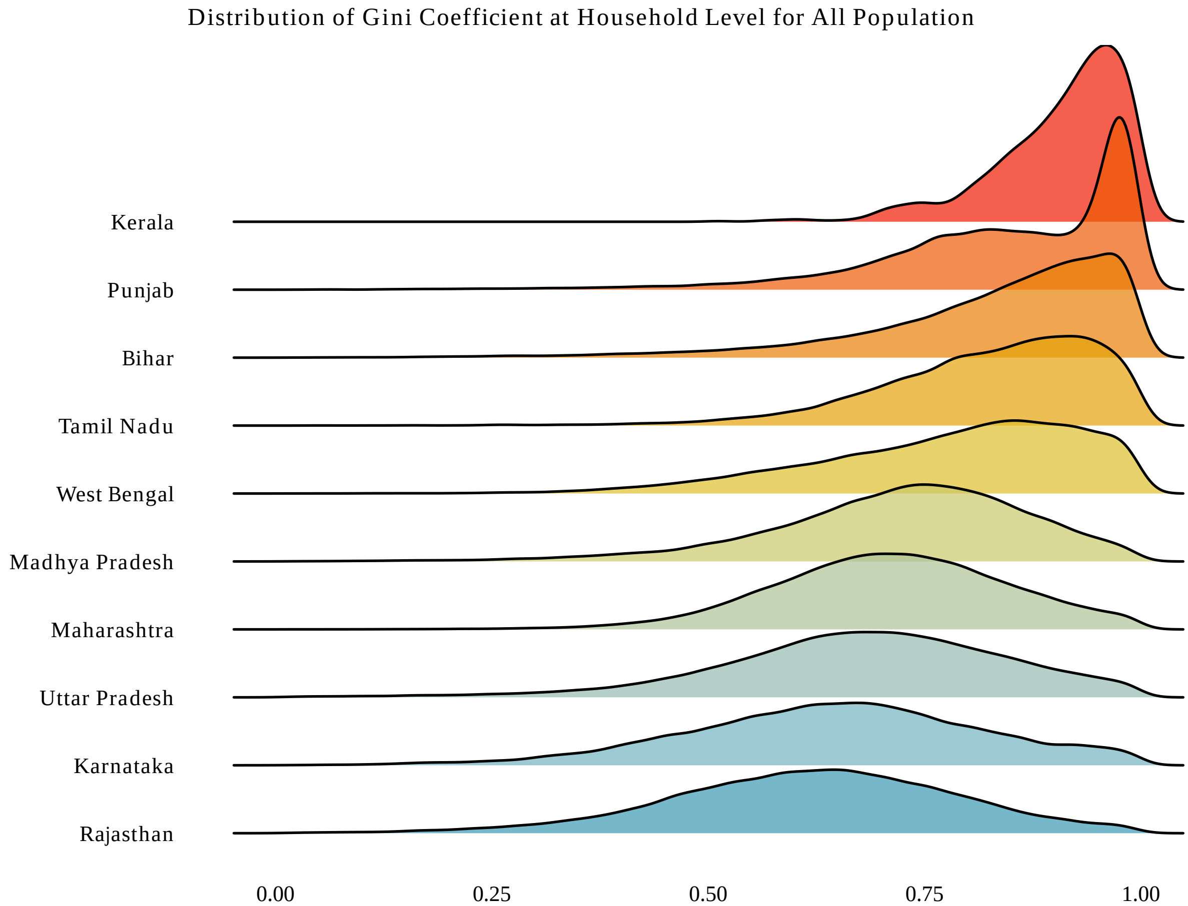
Tiwari, C., S. Goli, M. Z. Siddiqui, and P. S. Salve (2022): "Poverty, wealth inequality and financial inclusion among castes in Hindu and Muslim communities in Uttar Pradesh, India," *Journal of International Development*, 34, 1227–1255.

Varun, K. (2024): "Indirect rule and the unequal burden of persistence: Evidence from Hyderabad," *SSRN*.

Verghese, A. and E. Teitelbaum (2019): "Conquest and conflict: The colonial roots of Maoist violence in India," *Politics & Society*, 47, 55–86.

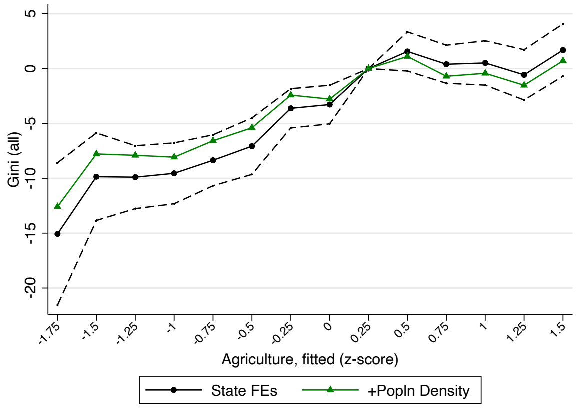
7 Figures

Figure 1: **Gini Distribution by State**

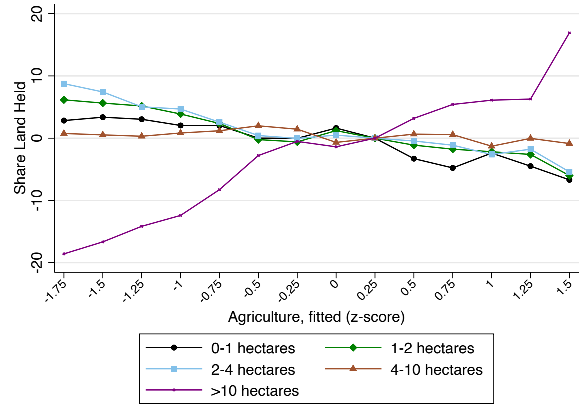


Notes: The figure plots the distribution of agricultural land gini including landlessness households in different states of India.

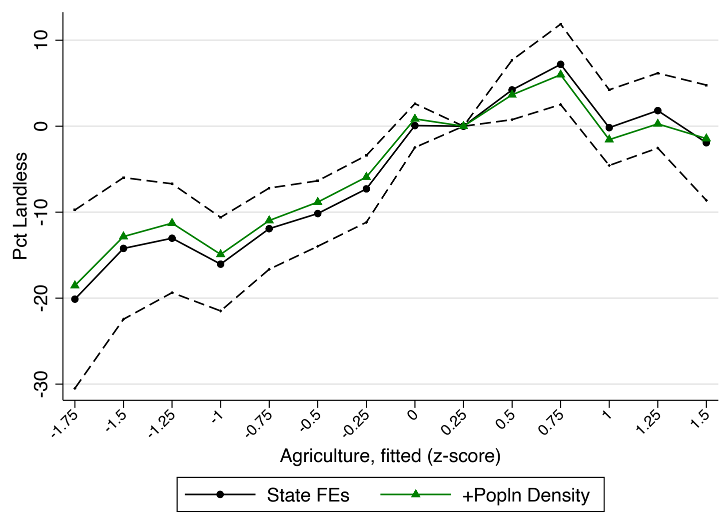
Figure 2: Agricultural Productivity and Inequality



(a) All-household



(b) Land ownership shares by type

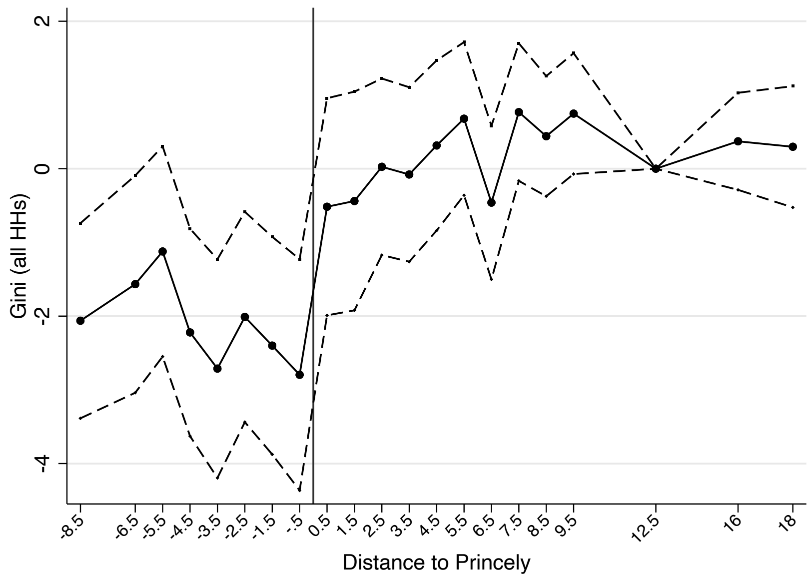


(c) Landlessness

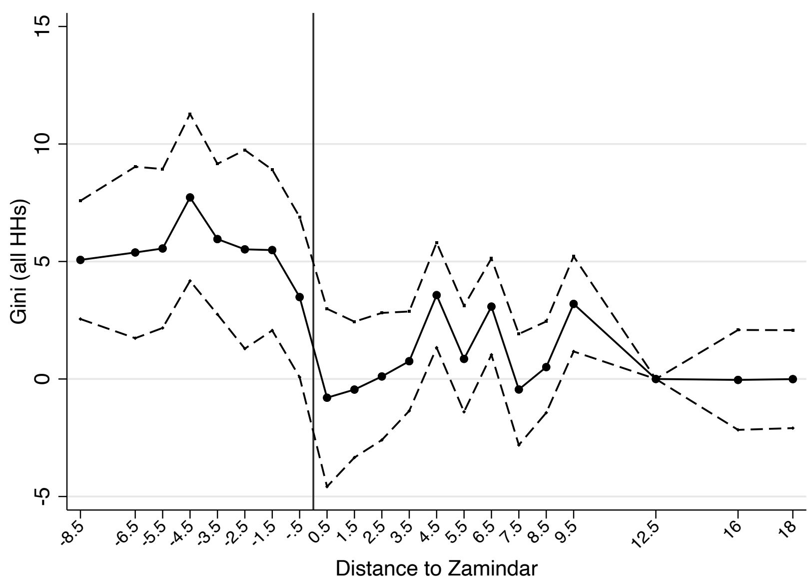
Note: Figure 2 plots coefficient estimates capturing the relationship between agricultural productivity (fitted values) and all-household land gini, landlessness, and the share of land owned by different landlord types, based on equation 1b. Standard errors are clustered at the district level. The sharp increase in land inequality as land productivity rises up to the 60th percentile is driven by a combination of increasing landlessness, and declining land shares of small and semi-medium landlords, and a pronounced rise in the land share of large landlords. Notably, the land share of medium-sized landlords remains largely unchanged throughout.

Figure 3: Distance to Institutions

(a) Princely

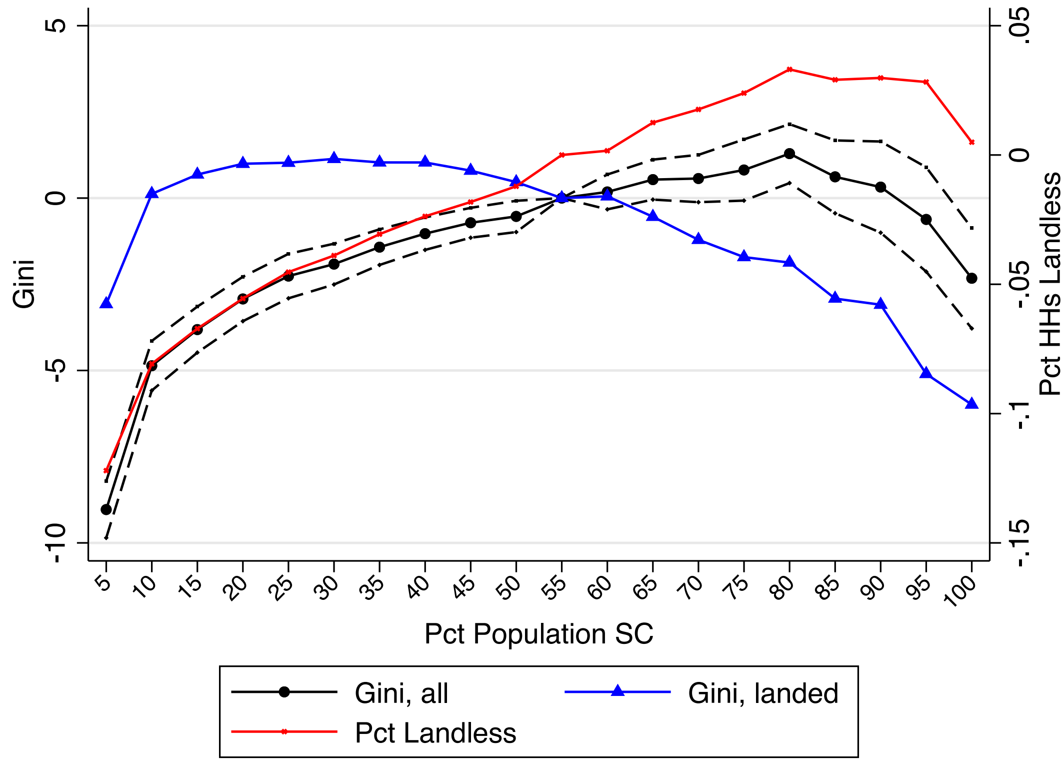


(b) *Zamindar*



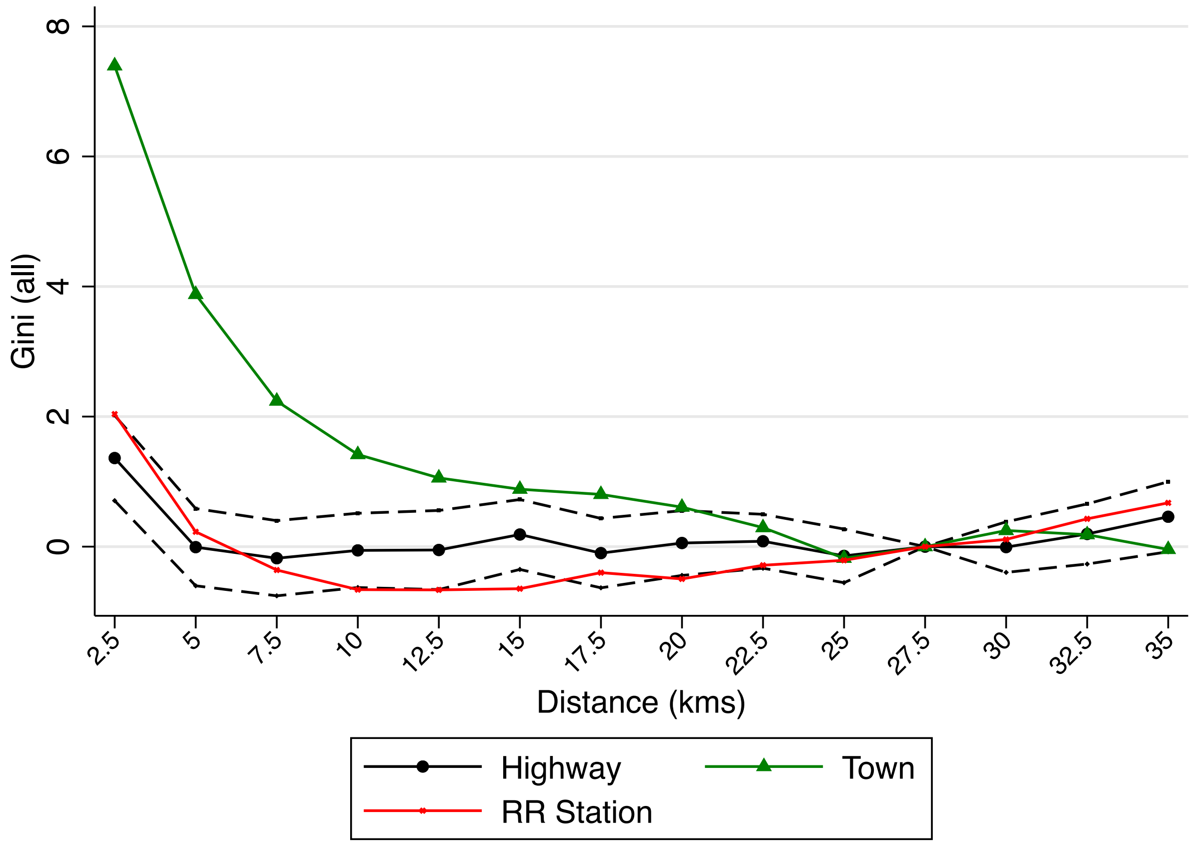
Note: Figure 3 plots the relationship between distance to *zamindar* areas and princely states and Gini for all households in top and bottom graphs respectively. The coefficients are following equation 6, where we control for vectors of agricultural ($\mathbf{A}_{vds}$), as well as 20-km boundary segment fixed effects. For princely state analysis we additionally include state fixed effects. For *zamindar* analysis we exclude state fixed effects and include the market variables ($\mathbf{M}_{vds}$), caste variables, and supplemental control variables. Standard errors are clustered at the boundary segment level. Negative distance denotes **inside** princely states or *zamindari* area and positive values denote **outside**. Princely states are associated with decreases in inequality, and *zamindar* areas with increases.

Figure 4: **Caste and Inequality**



Note: Figure 4 plots the relationship between percent of SC population in a village and three land inequality measures – all households Gini, landowner Gini, and share of landless households. Standard errors are clustered at district level. The relationship for all-household Gini is positive and largely linear in the share of the population being SC between 5–80%; is far steeper when the share increases from 5–10%; and is negative for values greater than 80%. This relationship is entirely driven by landlessness, with the relationship between the landed-household Gini and the SC population, in fact, being somewhat negative. The plotted coefficients ($\hat{\beta}_i$) are following equation 5, where we control full set of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$)-related variables, along with total village land area, and district fixed effects.

Figure 5: **Markets and Inequality**



Note: Figure 5 plots the relationship between the Gini (all households) and distance to: major highways, and railroad stations. For each market-related variables, the Gini coefficient increases substantially as proximity to the respective locations increases. Standard errors are clustered at district level. The increase is the largest for towns. For roads and railway stations the relationship is smaller, but still positive. Moreover, we see that inequality is elevated up to further distances from towns (10kms) than roads and railway stations (2.5kms). The plotted coefficients ($\hat{\beta}_i$) are following equation 6, where we control full set of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$)- related variables, along with total village land area, and district fixed effects.

8 Tables

Table 1: Inequality Correlates

Outcome: Gini Coefficients, All Households					
	(1)	(2)	(3)	(4)	(5)
Agriculture					
Altitude	-2.467***			-1.663**	-1.654**
	(0.906)			(0.728)	(0.725)
Ruggedness	-2.584**			-0.158	0.124
	(1.051)			(0.824)	(0.683)
Latitude	0.113			0.153	0.969***
	(0.178)			(0.176)	(0.264)
Growing Days	4.362***			2.966***	1.615*
	(1.056)			(1.126)	(0.899)
Temperature	-0.400			1.716	2.768**
	(1.829)			(1.692)	(1.234)
Precipitation	0.847			1.316***	1.799***
	(0.541)			(0.505)	(0.495)
Command Area	-0.931			-0.737	-0.351
	(0.682)			(0.590)	(0.430)
Distance River	-1.306***			-1.204***	-1.055***
	(0.354)			(0.306)	(0.202)
Alluvial Aquifer	1.427			1.127	1.595**
	(1.032)			(0.920)	(0.743)
Soil Suitability Index	-0.071***			-0.063***	-0.023**
	(0.014)			(0.013)	(0.010)
Market					
Distance Major Road		-1.649***		-1.559***	-1.085***
		(0.297)		(0.246)	(0.216)
Distance Town		-3.142***		-2.554***	-2.581***
		(0.555)		(0.376)	(0.299)
Distance RR Station		-2.167***		-0.870***	-0.874***
		(0.346)		(0.255)	(0.202)
Historic					
Princely State			-3.715***	0.298	-1.947*
			(1.075)	(1.031)	(1.044)
Zamindar			8.790***	6.740***	1.251
			(1.187)	(1.031)	(1.297)
Pct Popln SC			12.996***	12.961***	12.428***
			(1.225)	(1.123)	(0.848)
Pct Popln ST			-2.873*	0.114	0.784
			(1.543)	(1.378)	(1.154)
R-squared	0.113	0.059	0.109	0.181	0.253
N	276777	289261	281843	275037	275037
State Fixed Effects	No	No	No	No	Yes

Notes: This table presents the results of a regression of the Gini coefficient (for all households) on the full set of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$) covariates defined in village v, located in district d and state s. In Column (5) we include state fixed effects. Standard errors are clustered at district level.

Table 2: Shapley Decomposition

Factor	**Shapley Value** *estimate*	% *contribution*	**Shapley Value** *estimate*	% *contribution*	**Shapley Value** *estimate*	% *contribution*
Panel A: All Households						
Agriculture	0.068	37%	0.035	14%	0.006	2%
Market	0.036	20%	0.028	11%	0.018	5%
History	0.076	39%	0.029	12%	0.026	8%
Fixed Effects			0.161	64%	0.292	85%
R2	0.183		0.252		0.342	
Panel B: Landowning Households						
Agriculture	0.047	41%	0.017	13%	0.007	3%
Trade	0.004	3%	0.002	2%	0.002	1%
History	0.065	56%	0.017	13%	0.004	0%
Fixed Effects			0.095	73%	0.207	92%
R2	0.116		0.131		0.225	
Fixed Effects	None		State		District	

Notes: This table gives the Shapley decomposition values denoting the relative importance of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$) covariates in explaining the Gini (all households) land inequality in Panel A and Gini within landed community in Panel B.

Table 3: Impacts of Institutions on Land Inequality Measures

	Princely State			*Zamindar* Tenure		
Outcomes:	(1)	(2)	(3)	(4)	(5)	(6)
Gini, all	-1.206	-2.483***	-1.322	7.076***	-0.882	8.417***
	(0.894)	(0.777)	(0.847)	(1.048)	(1.531)	(1.041)
Gini, landowners	0.316	-0.238	0.784	5.712***	-1.409	6.353***
	(0.594)	(0.674)	(0.615)	(1.131)	(0.917)	(1.143)
Pct Landless	-1.781	-4.097***	-2.462	9.371***	0.164	11.528***
	(1.582)	(1.487)	(1.551)	(1.989)	(2.741)	(1.971)
Pct Land held by those with						
<2 hectares	0.010	2.839*	-0.226	-6.248***	2.404	-6.675***
	(1.317)	(1.478)	(1.290)	(1.734)	(3.517)	(1.705)
$2-4$ hectares	-1.121**	0.040	-1.078**	-2.434***	1.725	-2.788***
	(0.513)	(0.633)	(0.512)	(0.574)	(1.048)	(0.584)
$4-10$ hectares	-2.108***	-1.517**	-2.390***	-1.875**	1.165	-2.211***
	(0.597)	(0.664)	(0.593)	(0.830)	(1.423)	(0.841)
>10 hectares	4.194***	0.851	3.639***	11.174***	-0.930	11.393***
	(1.191)	(1.287)	(1.183)	(1.632)	(2.767)	(1.689)
largest farm	1.417***	0.326	1.005**	5.174***	1.764***	5.220***
	(0.507)	(0.496)	(0.474)	(0.945)	(0.647)	(0.938)
Landlord Village	2.162***	0.881	1.684**	7.732***	2.014**	7.993***
($>30\%$ land)	(0.773)	(0.853)	(0.745)	(1.610)	(0.913)	(1.594)
Landlord Village	0.880**	0.314	0.745**	4.578***	0.692	4.604***
($>50\%$ land)	(0.372)	(0.370)	(0.353)	(0.990)	(0.430)	(0.957)
State FEs	No	Yes	No	No	Yes	No
Supplemental Controls	No	No	Yes	No	No	Yes

Notes: This table gives the impact of princely state and *zamindar* institutions on the indicated measures of land distribution. Each coefficient ($\hat{\Phi}_i$) comes from a separate regression following equation 3, including the full vector of agriculture, market, and historic-related variables. Columns (1)–(3) include all villages; columns (4)–(6) exclude villages in princely states. Columns (2) and (5) include state fixed effects. Columns (3) and (6) exclude state fixed effects, but include additional controls, including: share of the population outside of agriculture, (ln) light density in 2012, an indicator for paved road, (ln) number of workers in firms with $10-50$ employees, and (ln) number of workers in firms with >50 employees. Landlord Village is an indicator which takes a value of 1 if the share of land owned by the largest landholder is greater than 30% or 50%, respectively. Standard errors are clustered at district level.

Table 4: Princely State and Land Inequality: Border Discontinuity Design

	Land Gini, All Households					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline						
Princely State	-2.736**	-2.686**	-3.166***	-3.594***	-2.508***	-2.651***
	(1.171)	(1.192)	(0.983)	(0.973)	(0.720)	(0.645)
R-squared	0.162	0.157	0.220	0.223	0.342	0.361
N	45793	23011	45793	23011	45775	22996
Panel B: Agr Controls						
Princely State	-2.150**	-2.513**	-1.996***	-2.670***	-2.233***	-2.458***
	(0.990)	(0.967)	(0.761)	(0.730)	(0.713)	(0.650)
R-squared	0.198	0.196	0.245	0.249	0.351	0.371
N	43643	21861	43643	21861	43623	21844
Panel C: Agr + Historic (Caste) Controls						
Princely State	-1.964*	-2.254**	-1.899**	-2.483***	-2.204***	-2.400***
	(1.054)	(1.012)	(0.754)	(0.717)	(0.687)	(0.661)
R-squared	0.227	0.225	0.278	0.284	0.380	0.400
N	43541	21799	43541	21799	43520	21782
Panel D: Agr + Historic (Caste) + Running Variable						
Princely State	-2.211**	-2.282**	-2.526***	-2.616***	-2.320***	-2.044***
	(1.064)	(0.999)	(0.810)	(0.756)	(0.707)	(0.784)
R-squared	0.227	0.225	0.278	0.284	0.380	0.400
N	43541	21799	43541	21799	43520	21782
BW (kms)	20	10	20	10	20	10
State FEs	Yes	Yes	Yes	Yes	Yes	Yes
Princely State FEs			Yes	Yes		
Segment FEs					Yes	Yes

Notes: This table shows the results of a regression of all-household Gini on an indicator variables for villages located within the princely state. Panel A excludes control variables; Panel B includes the vector of agricultural variables ($\mathbf{A}_{vds}$); Panel C includes both agricultural and historic (caste) variables; and Panel D further adds the running variable (distance to the princely-state border) and its treatment interaction to account for smooth spatial trends across the boundary. Standard errors are clustered at district level. The coefficients are relatively stable across specifications with princely states having approximately 2–3 percentage points lower land inequality.

Table 5: *Zamindar*, UP-Bihar Boundary (BDD)

	Land Gini, All Households							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Zamindar	5.596**	3.690**	3.835***	2.916**	3.821***	2.970**	4.486***	3.749***
	(1.996)	(1.382)	(1.196)	(1.087)	(1.171)	(1.025)	(1.187)	(1.171)
R-squared	0.207	0.192	0.230	0.206	0.253	0.234	0.295	0.275
N	8252	4374	8157	4314	8145	4307	6649	3521
BW (kms)	20	10	20	10	20	10	20	10
Segement FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls:								
Agriculture			Yes	Yes	Yes	Yes	Yes	Yes
SC/ST					Yes	Yes	Yes	Yes
Supplemental							Yes	Yes

Notes: This table shows the results of a regression of all-household Gini on an indicator variables for villages located within the a *zamindar* area. The sample is restricted to villages along the border of the two largest states Uttar Pradesh-Bihar. Columns (3)-(8) include agriculture variables; and Columns (5)-(8) include SC/ST variables. Columns (7) and (8) additionally include the supplemental control variables, which include: (ln) light density; share of the male population working outside agriculture; the presence of a paved road; and (ln) of workers in firms with 50–100 employees and (ln) of workers in firms with >100 employees. Standard errors are clustered at district level. We see that the Gini coefficient in *zamindar* areas is approximately 3–5 percentage points higher significant at 1% level.

Table 6: Structural Transformation and Inequality

Outcome: Gini Coefficient, All	(1)	(2)	(3)	(4)
High Manu	7.182***	5.922***	4.449***	2.978***
	(1.199)	(1.190)	(0.729)	(0.693)
Fitted Agri	3.336***	2.702***	2.544***	1.012**
	(0.677)	(0.593)	(0.454)	(0.428)
Princely	-2.850***	-2.744***	-0.366	-0.204
	(0.962)	(0.995)	(0.398)	(0.391)
Zamindar	1.123	1.467	-0.752	-0.893
	(1.482)	(1.340)	(1.846)	(1.617)
SC	11.508***	11.560***	12.020***	11.219***
	(0.993)	(1.003)	(0.803)	(0.709)
ST	0.263	1.037	-0.241	1.801**
	(1.355)	(1.467)	(0.916)	(0.868)
High Manu X				
Fitted Agri	-1.369*	-1.805**	-1.381**	-1.623**
	(0.771)	(0.769)	(0.695)	(0.758)
Princely	1.104	1.226	-0.509	0.333
	(1.475)	(1.505)	(0.972)	(1.038)
Zamindar	-0.957	-1.851	-0.483	-0.440
	(2.355)	(2.123)	(0.979)	(0.951)
SC	0.082	-0.129	2.708	2.777
	(2.208)	(2.148)	(2.212)	(1.975)
ST	-0.753	0.749	1.361	2.196
	(2.690)	(2.772)	(2.412)	(2.244)
R-squared	0.232	0.289	0.340	0.393
N	260586	260499	260586	260499
Fixed Effects	State		District	
Controls:				
Ln Popln Density	No	Yes	No	Yes
Town Prox w/Interactions	Yes	Yes	Yes	Yes

Notes: This table presents the results of a regression of the (all-household) Gini coefficient on the indicated variables. *High Manu* is a dummy (=1), defined at sub-district level in which the share of the working population outside of agriculture is more than one standard deviation above the mean, and *Fitted Agri* is the z-score of the fitted value of NDVI using the agricultural variables – $\widehat{\text{EVI}}_{vds}$ from equation 1a. A control is also included for town proximity and its interaction with the agriculture and historic variables; and in columns (2) and (4) a control is included for ln population density. Columns (1)–(2) include state fixed effects; and column (3)–(4) district fixed effects. Standard errors are clustered at district level. In high-manufacturing areas, the influence of agricultural-productivity on inequality is reduced by 50–100%. In contrast, there is no decline in the role of historic factors in high-manufacturing areas–all the interaction coefficients are insignificant.

Table 7: Landlords and Public Goods

	Gini All HHs		Gini Landowning HHs			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Govt Primary School						
Landlord	-0.104***	-0.091***	-0.123***	-0.115***	-0.122***	-0.115***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R-squared	0.289	0.290	0.292	0.293	0.293	0.293
N	260356	260356	260356	260356	260340	260340
Panel B: Paved Road						
Landlord	-0.027***	-0.018***	-0.041***	-0.031***	-0.041***	-0.032***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
R-squared	0.313	0.313	0.314	0.314	0.315	0.315
N	260055	260055	260055	260055	260040	260040
Panel C: Primary Health Center						
Landlord	-0.014***	-0.018***	-0.020***	-0.019***	-0.020***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.057	0.058	0.058	0.059	0.063	0.063
N	259841	259841	259841	259841	259826	259826
Panel D: Sanitation Campaign						
Landlord	-0.007**	-0.007**	-0.009***	-0.006**	-0.009***	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
R-squared	0.442	0.442	0.442	0.442	0.442	0.442
N	259879	259879	259879	259879	259864	259864
Gini Sample	All HHs		Landowning HHs			
Gini Polynomial Control	Linear	Cubic	Linear	Cubic	Linear	Cubic
Pct Landless Control			Yes	Yes	Yes	Yes
Control Non-Agr Labor					Yes	Yes

Notes: This table presents the results of a regression of the indicated public good on an indicator for whether there is a major landlord (one individual owning more than 30% of the land). All regressions control for full set of agriculture, market and historic-related covariates, social fractionalization index, population density in logs, and district fixed effects. Standard errors are clustered at district level. There is a decline in all public goods when the village is dominated by a landlord, with the strongest decline in the presence of government public schools. These results are robust to the inclusion of either a linear or upto cubic in Gini (Columns (2) and (4)), or controlling the share of population in non-agricultural sector (Columns (5)- (6)).

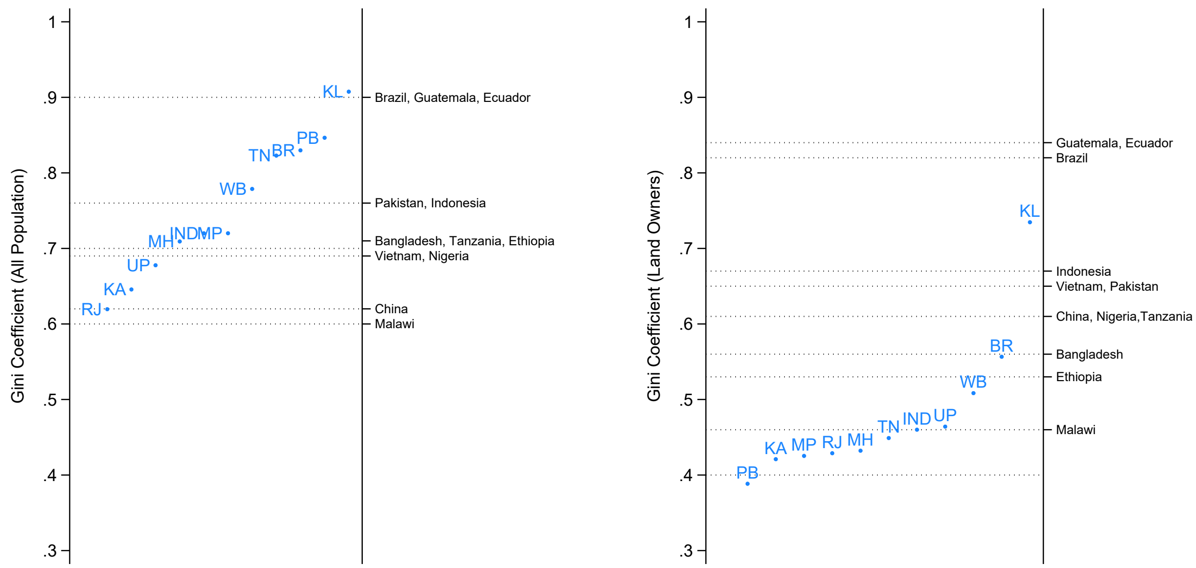
Appendix: Land Inequality in India: Nature, History, and Markets

March 1, 2026

Nitin Kumar Bharti - University of Western Australia.
David Blakeslee - New York University Abu Dhabi.
Samreen Malik - New York University Abu Dhabi.

Appendix A

Figure A1: Land Area Inequality in Comparison with Other Countries



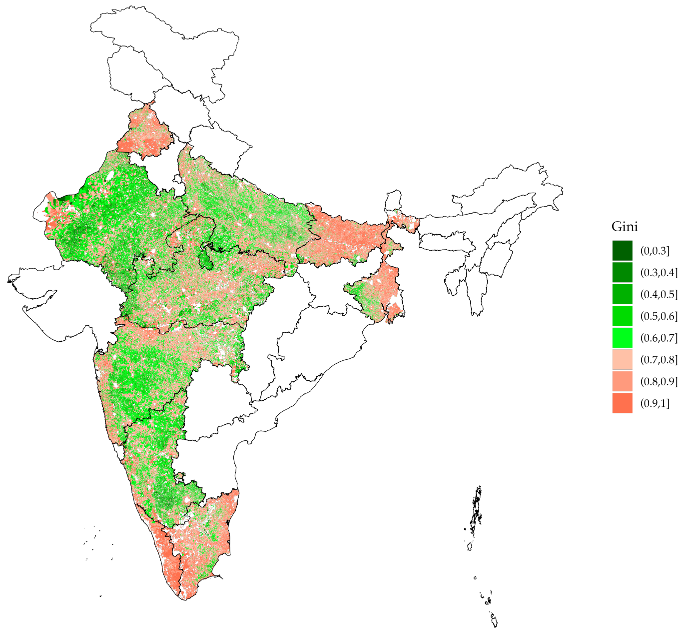
(a) Gini (All population)

(b) Gini (within landowners)

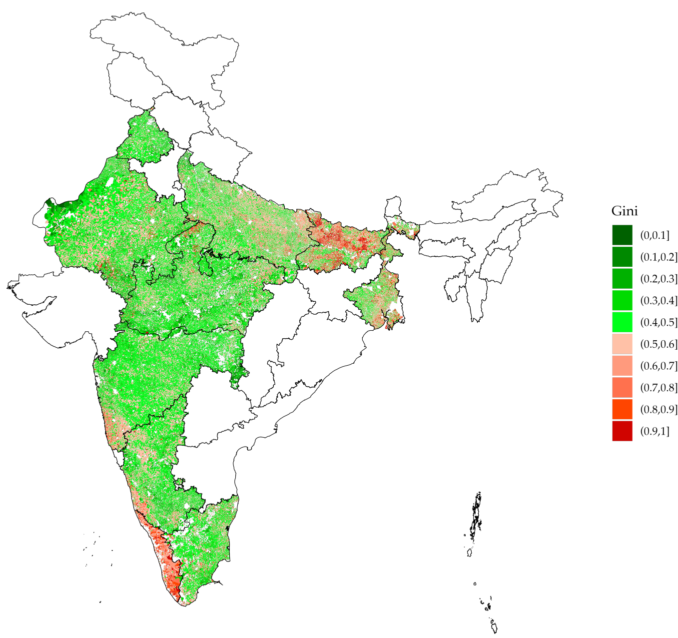
Notes: The figure represents the average village-level land inequality (Gini) among the total population and among landowners (excluding the landless) in ten large Indian states, compared with other countries (Bauluz et al., 2025). RJ—Rajasthan; KA—Karnataka; UP—Uttar Pradesh; MH—Maharashtra; MP—Madhya Pradesh; WB—West Bengal; TN—Tamil Nadu; BR—Bihar; PB—Punjab; KL—Kerala.

Figure A2: Land inequality distribution map

(a) **Gini, all households**

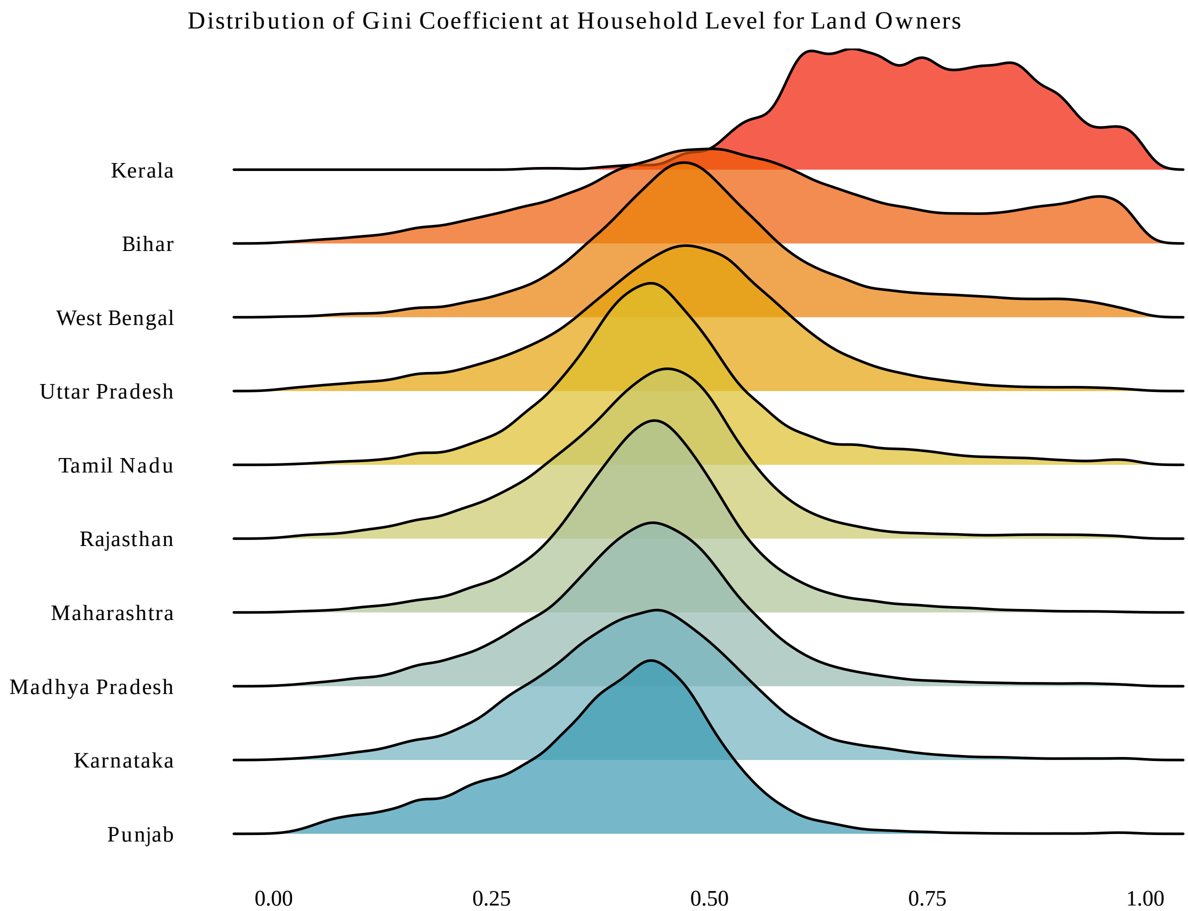


(b) **Gini, landowning households**



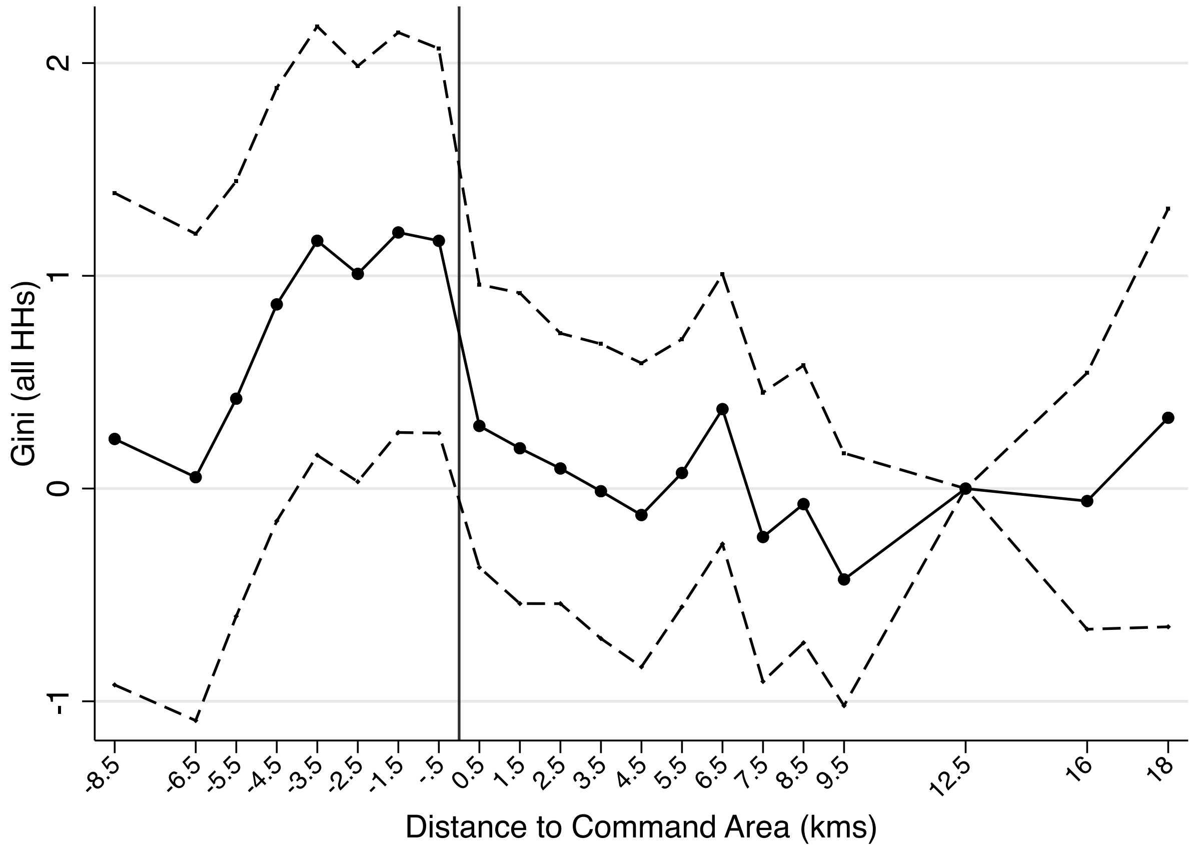
Notes: The figure plots the distribution of agricultural land gini total and gini within landed households in India.

Figure A3: **Gini Distribution, landowning households**



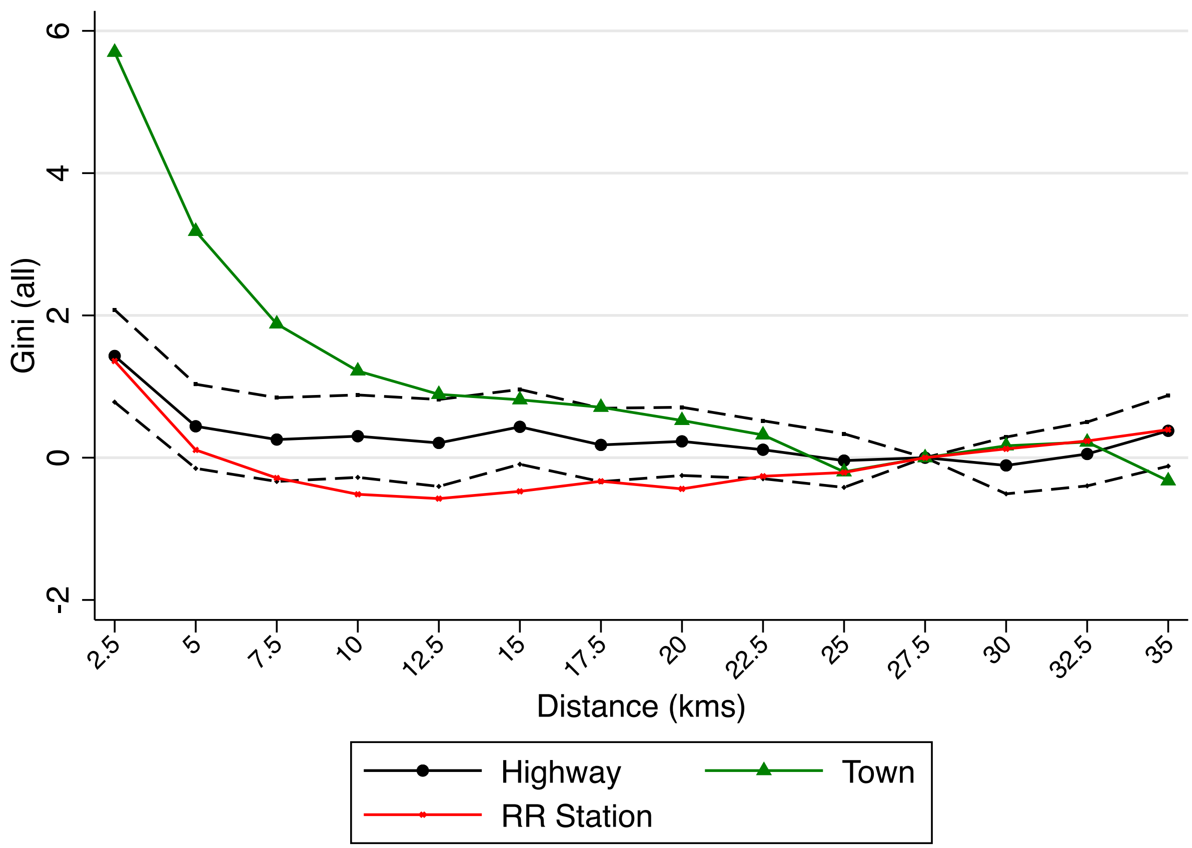
Notes: The figure plots the distribution of agricultural land gini among landowing households in different states of India.

Figure A4: **Irrigation and Inequality**



Notes: The figure plots the coefficient estimates with outcome variable as all-household Gini against the distance from the border of the irrigation scheme, also called "command area", following equation 2, controling for agriculture agricultural variables ($\mathbf{A}_{vds}$), and including district and 10km boundary segment fixed effects. The negative value on the x-axis denotes the presence of village within the command area. Standard errors are clustered at command area project level. Land inequality is on average 1 pp higher in villages within the command areas.

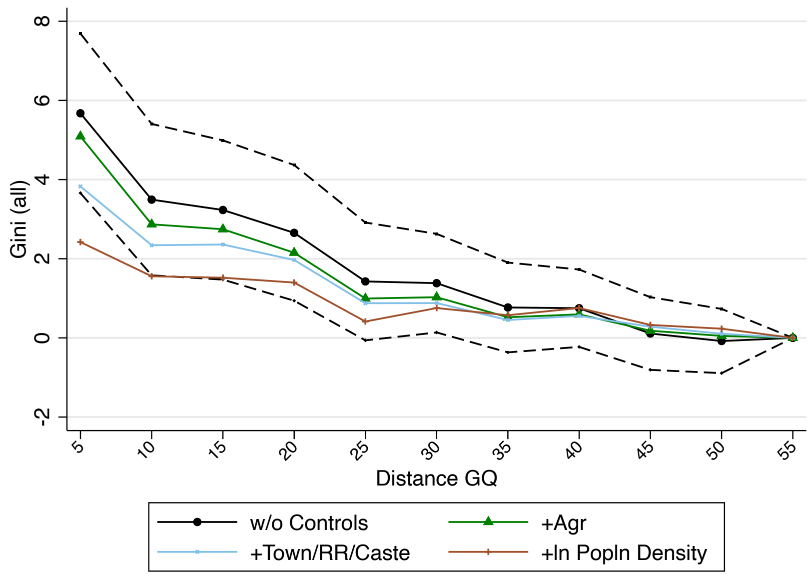
Figure A5: **Trade and Inequality w/Population Control**



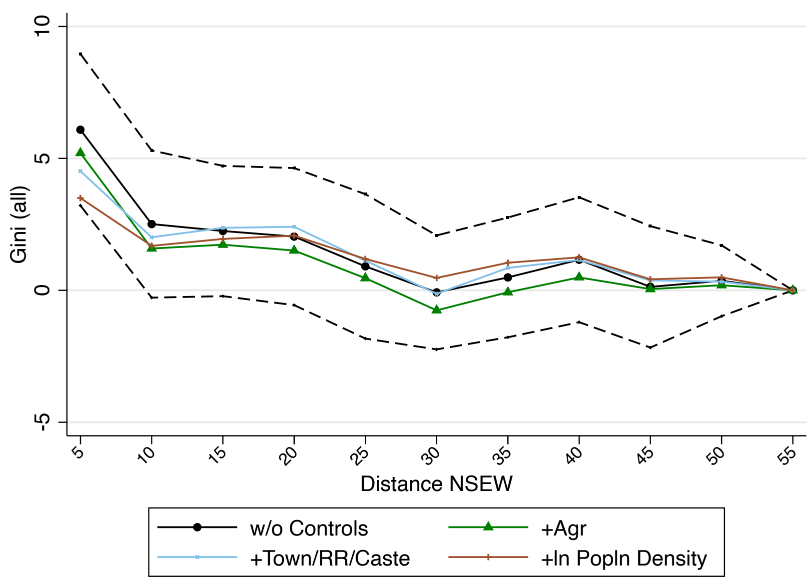
Note: Figure A5 plots the relationship between the Gini (all households) and distance to: major highways, and railroad stations. This figure is similar to Figure 5, giving the coefficients from equation 6, while including an additional control for the population density in logs. Standard errors are clustered at the district level. The results are largely unchanged, though the impacts are somewhat reduced.

Figure A6: Distance to Important National Highways

(a) Golden Quadrilateral



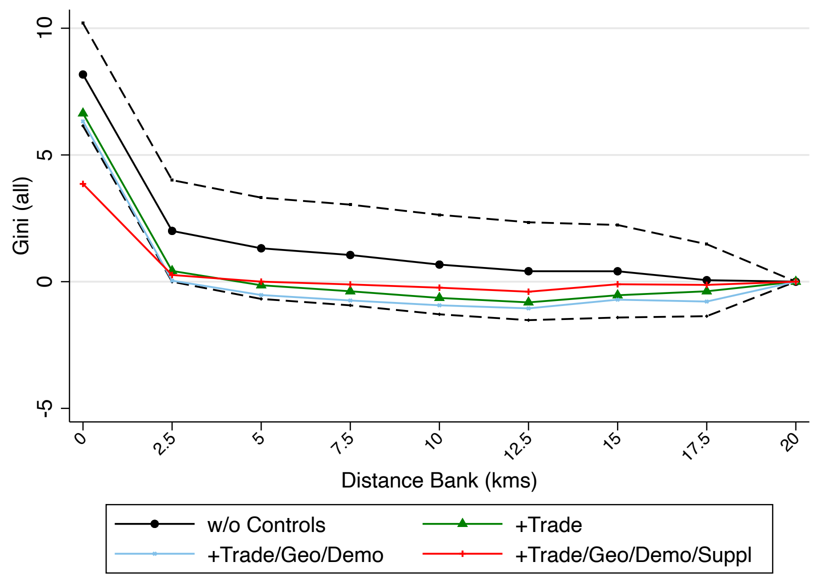
(b) North–South–East–West



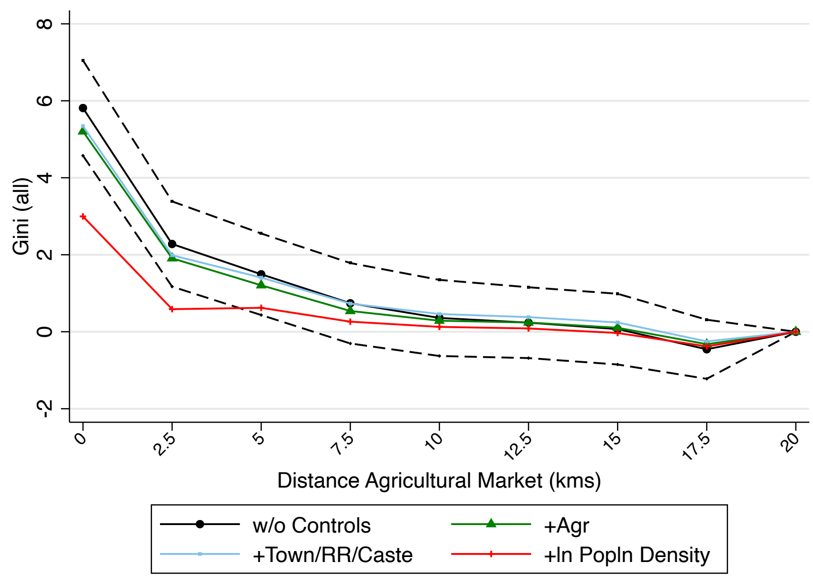
Note: Figure A6 plots the relationship between land inequality (Gini, all households) and distance to major national highways: the Golden Quadrilateral (panel (a)) and the North–South–East–West highway system (panel (b)). Coefficient estimates are based on equation 6, controlling for vectors of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$) variables, as well as village land area, population density, and district fixed effects. Standard errors are clustered at the district level. For the Golden Quadrilateral—long-established major highways—land inequality increases up to 25 km from the highway, reaching nearly 2 percentage points in the immediate vicinity. In contrast, the impact of the NSEW highways is less than half as large, with no evidence of increased inequality beyond 5 km.

Figure A7: Distance to Banks and Agricultural Markets

(a) Bank

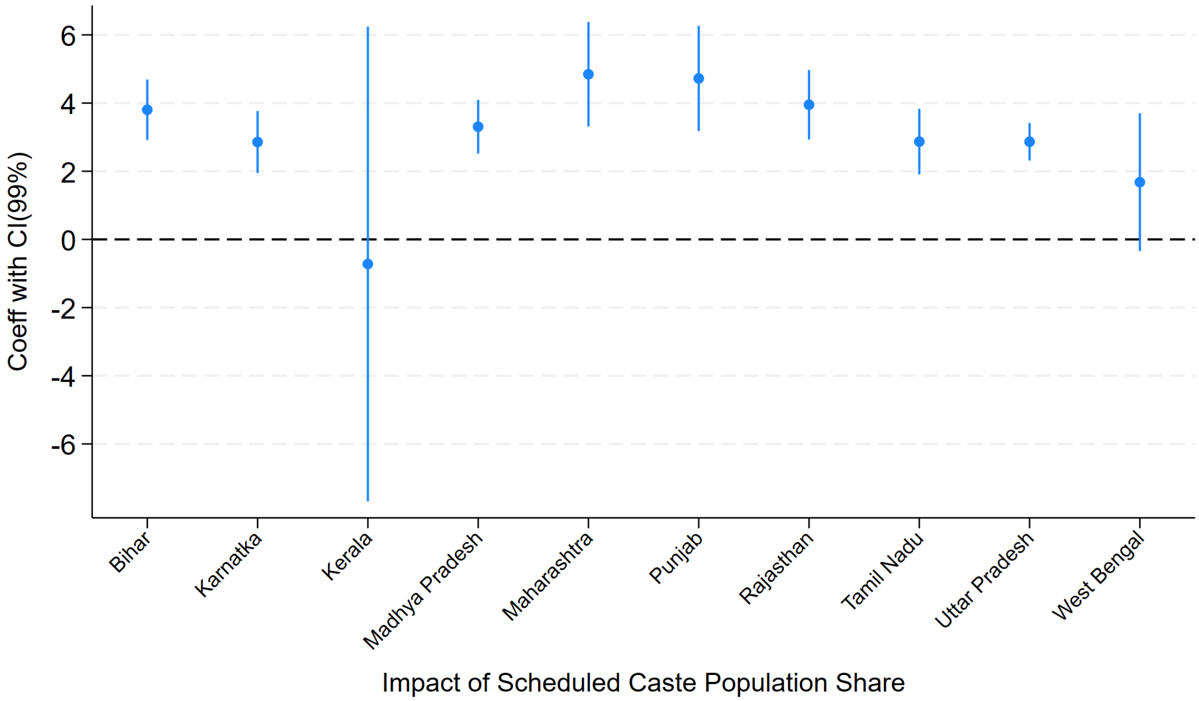


(b) Mandi



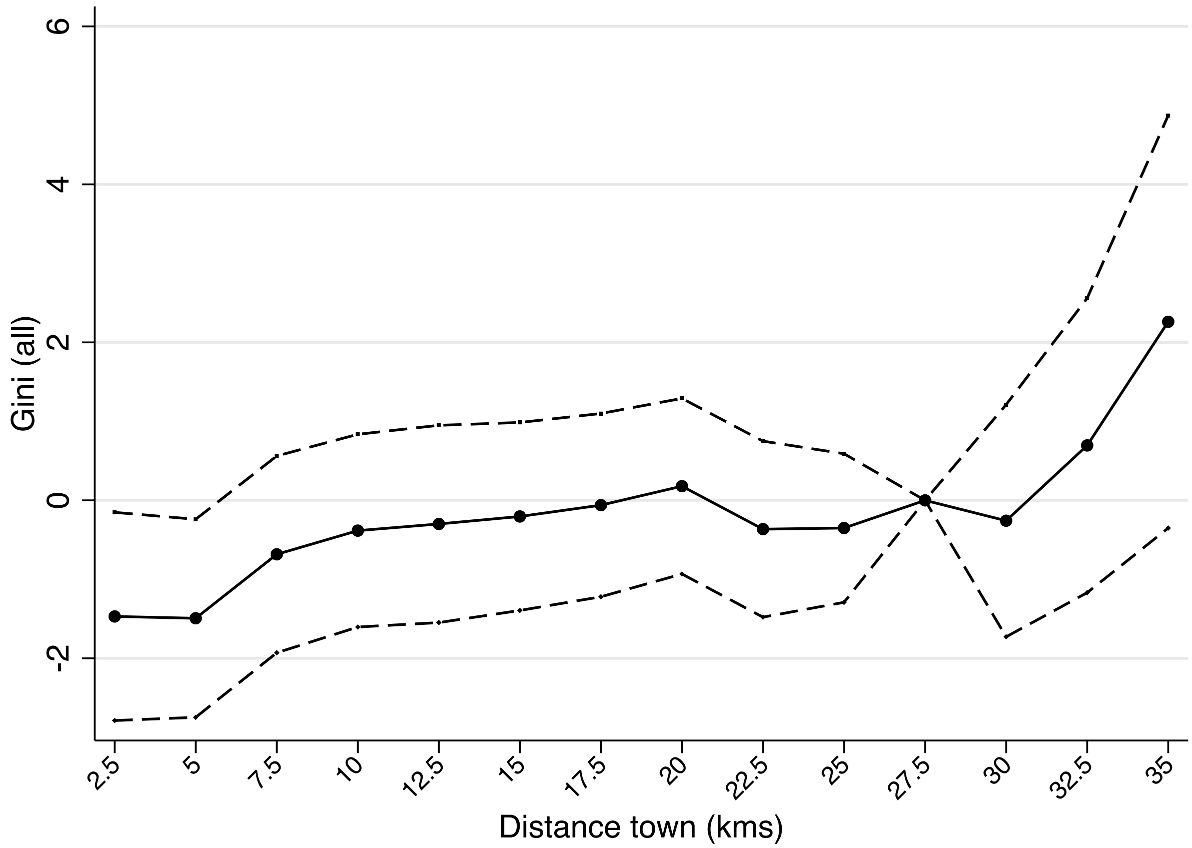
Note: Figure A7 plots the relationship between distance to banks and agricultural markets (mandis) and Gini for all households in top and bottom graphs respectively. The coefficients are following equation 6, where we control for vectors of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$)- related variables, along with village land area, population density, and district fixed effects. Standard errors are clustered at district level. Both bank presence and agricultural markets are associated with greater land inequality, with little evidence for spillovers to villages even very close to the facilities.

Figure A8: **Impacts of SC Population Share on Gini (all households)**



Note: Figure A8 plots the relationship between the Gini (all households) and standardized Scheduled Caste population share (by state's share and standard deviation), separately by state. The coefficients for regressions run separately for individual states. District fixed effects are included, as well as the agricultural and demographic variables, and (ln) population density. Standard errors are clustered at district level. For most states, changing the SC population by 1 standard deviation, increases the land gini between 3–5 and is statistically significant at 1% level. One notable exception to this pattern is the states of Kerala and West Bengal, which were long governed by left-wing parties, and in which post-independence land reforms are generally considered to have been the most successful: in these states, the relationship between SC population share and land inequality is substantially smaller and is statistically insignificant.

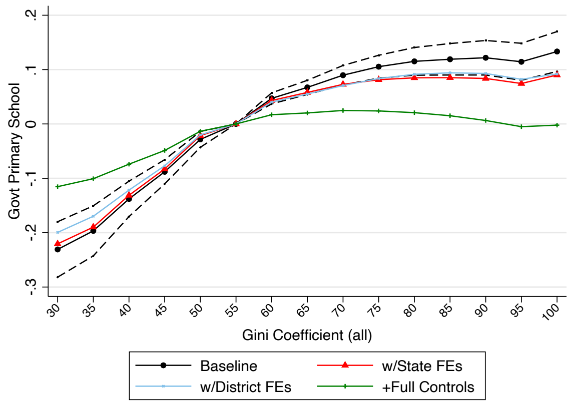
Figure A9: **Agricultural Productivity, Distance to Town, and Inequality**



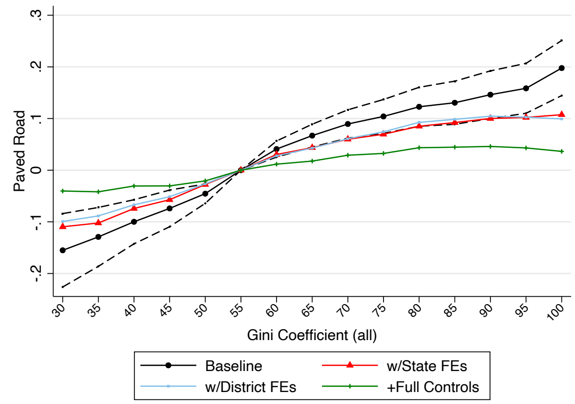
Note: Figure A9 plots coefficients for the interaction terms of (fitted) agricultural output and the indicated distance indicators. The fitted agriculture is the z-score of the fitted value of NDVI using the agricultural-related variables – $\widehat{\text{EVI}}_{vds}$ from equation 1a. The gini for household is then regressed on distance to town, $\widehat{\text{EVI}}_{vds}$, their interaction, alongwith full vector of our usual controls, with district fixed effects. Standard error is clustered at district level. This graphs shows the decline in the relationship between agriculture and inequality in close proximity to towns, as well as a sharp increase at distances greater than 30kms.

Figure A10: Gini and Public Goods

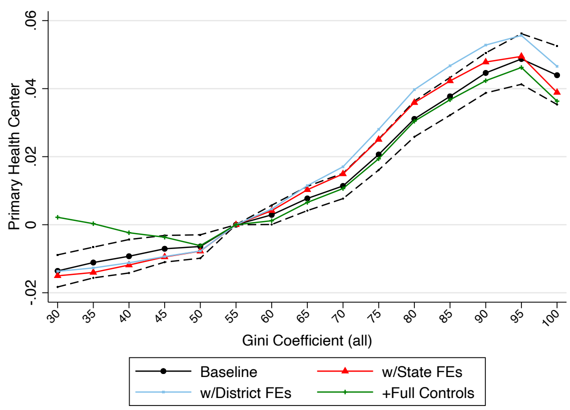
(a) Govt Primary School



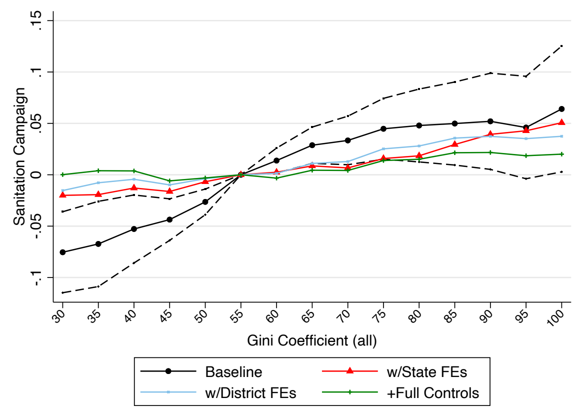
(b) Paved Road



(c) Primary Health Center



(d) Sanitation Campaign



Note: Figure A10 plots the relationship between (all-household) Gini and the presence of the indicated public goods. Results are shown with and without control variables. The full controls include the agriculture, market, and history variables, as well as ln population density and social fractionalization index (following Banerjee and Somanathan, 2007). Standard errors are clustered at district level. While the raw correlation is generally upwards sloping and concave, the inclusion of controls leads to an inverted-U relationship for some outcomes. This suggests that intermediate levels of inequality may promote the provision of public goods, but that extreme inequality is detrimental to public goods provision.

Table A1: Summary Statistics

	Mean	Percentile				
		p10	p25	p50	p75	p90
	(1)	(2)	(3)	(4)	(5)	(6)
Gini, All HHs	0.711	0.494	0.607	0.723	0.836	0.922
Gini, Landowners	0.459	0.281	0.371	0.452	0.533	0.639
Pct HHs Landless	0.464	0.124	0.267	0.459	0.656	0.814
Avg Landholding, All HHs (hectares)	2.574	0.173	0.346	0.667	1.280	2.279
Avg Landholding, Landowners (hectares)	6.158	0.488	0.810	1.427	2.445	4.049
Pct Land in Holdings of:						
0-1 hectares	0.289	0.037	0.092	0.220	0.428	0.651
1-2 hectares	0.486	0.215	0.355	0.493	0.618	0.737
2-4 hectares	0.210	0.053	0.113	0.192	0.278	0.375
4-10 hectares	0.284	0.086	0.172	0.275	0.375	0.476
$>$10 hectares	0.377	0.086	0.167	0.325	0.544	0.764
Top 10% land share	0.443	0.294	0.343	0.406	0.496	0.645
Top 5% land share	0.322	0.182	0.221	0.275	0.361	0.526
Top 1% land share	0.181	0.061	0.082	0.118	0.199	0.385
Top Landholder land share	0.124	0.027	0.042	0.072	0.137	0.281
Major Landlord Present	0.038					
Princely State	0.326					
Zamindar	0.225					
Distance Major Road (kms)	19.607	1.347	5.266	13.718	27.701	46.426
Distance Town (kms)	11.270	3.807	6.149	9.769	14.468	19.933
Distance RR Station (kms)	13.874	2.789	5.077	10.155	18.843	30.488
Pct Popln SC	0.203	0.000	0.045	0.157	0.298	0.466
Pct Popln ST	0.109	0.000	0.000	0.000	0.052	0.431

Notes: This table gives summary statistics. Column (1) is the village mean for the indicated variables. Columns (2)-(6) give the values at p10, p25, p50 (median), p75 and p90 percentiles respectively.

Table A2: Village-level summary statistics

	Gini		Pct HHs	Pct Land				Major
	All	Landed	Landless	Top 10%	Top 5%	Top 1%	Top Owner	Landlord Present
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bihar	0.828	0.553	0.594	0.561	0.451	0.306	0.201	0.120
Karnataka	0.646	0.425	0.386	0.412	0.293	0.154	0.098	0.030
Kerala	0.900	0.717	0.654	0.652	0.546	0.347	0.200	0.048
Madhya Pradesh	0.715	0.419	0.507	0.427	0.313	0.185	0.116	0.044
Maharashtra	0.709	0.431	0.481	0.408	0.288	0.147	0.088	0.022
Punjab	0.835	0.364	0.728	0.405	0.295	0.185	0.119	0.108
Rajasthan	0.621	0.424	0.340	0.419	0.300	0.166	0.114	0.034
Tamil Nadu	0.809	0.436	0.665	0.436	0.317	0.177	0.091	0.034
Uttar Pradesh	0.675	0.455	0.394	0.460	0.336	0.190	0.073	0.039
West Bengal	0.777	0.493	0.555	0.480	0.364	0.226	0.103	0.064

Notes: This table provides the average of village-level inequality statistics. The inequality measures are Gini for all households in Column (1), Gini within landed households in Column (2) and share of landless households in Column (3). Columns (4)-(6) provides the share of land owned (of total land) by the top 10%, top 5% and top 1% of the households by land ownership. Column 7 captures the share of total land owned by the household with the most agricultural land in a village. Column 8 depicts the share of villages with a major landlord, which is defined as one household owning more than 50% of the land.

Table A3: Inequality Correlates, Landholdholders

Outcome: Gini Coefficients Landholders					
	(1)	(2)	(3)	(4)	(5)
Agriculture					
Altitude	-0.646			-1.148	-0.212
	(0.724)			(0.697)	(0.737)
Ruggedness	-4.369***			-1.094	-1.357**
	(0.740)			(0.706)	(0.652)
Latitude	-0.216*			-0.081	-0.448*
	(0.128)			(0.118)	(0.237)
Growing Days	-0.352			-0.392	-0.112
	(1.017)			(0.931)	(0.885)
Temperature	-1.577			0.580	0.460
	(1.307)			(1.219)	(1.168)
Precipitation	2.203***			2.005***	1.550***
	(0.720)			(0.630)	(0.566)
Command Area	-0.307			0.078	0.395
	(0.539)			(0.477)	(0.470)
Distance River	-0.172			-0.208	-0.136
	(0.256)			(0.237)	(0.205)
Alluvial Aquifer	2.582***			2.152***	2.026***
	(0.648)			(0.634)	(0.620)
Soil Suitability Index	-0.009			0.004	-0.011
	(0.009)			(0.009)	(0.009)
Market					
Distance Major Road		-0.680***		-0.413**	-0.270
		(0.221)		(0.185)	(0.179)
Distance Town		0.089		0.670**	0.229
		(0.381)		(0.288)	(0.252)
Distance RR Station		-2.213***		-0.658***	-0.680***
		(0.321)		(0.221)	(0.188)
Historic					
Princely State			-1.769***	1.522**	-1.024
			(0.603)	(0.645)	(0.799)
Zamindar			6.272***	4.938***	-1.817**
			(1.170)	(1.113)	(0.893)
Pct Popln SC			-0.964	-1.326*	0.536
			(0.773)	(0.720)	(0.541)
Pct Popln ST			-11.522***	-8.554***	-7.615***
			(0.961)	(0.784)	(0.673)
R-squared	0.076	0.019	0.074	0.103	0.131
N	276777	289262	281844	275037	275037
State Fixed Effects	No	No	No	No	Yes

Notes: This table presents the results of a regression of the Gini coefficient (including only landholders) on the indicated variables. Columns (1)–(5) do no included fixed effects; Column (6) includes state fixed effects. Standard errors are clustered at district level.

Table A4: Command Areas, Agriculture, and Inequality

	Gini		EVI	
	(1)	(2)	(3)	(4)
Panel A: w/o Controls				
Command Area	0.644*	0.922***	4.103***	3.864***
	(0.369)	(0.348)	(0.840)	(0.851)
R-squared	0.421	0.437	0.810	0.829
N	101824	65588	155351	102126
Panel B: Agr Controls				
Command Area	0.626*	0.916**	2.915***	2.707***
	(0.365)	(0.363)	(0.770)	(0.791)
R-squared	0.425	0.440	0.821	0.839
N	98332	63304	150790	99056
Panel C: Agr + Historic				
Command Area	0.519	0.839**	2.313***	2.182***
	(0.359)	(0.354)	(0.724)	(0.745)
R-squared	0.444	0.460	0.829	0.848
N	97283	62601	141846	92719
Panel D: Agr + Historic + Running Variable				
Command Area	0.925**	1.083**	1.704**	1.818***
	(0.403)	(0.449)	(0.780)	(0.626)
R-squared	0.444	0.460	0.830	0.848
N	97283	62601	141846	92719
Control Mean	72.307	72.560	15.867	16.605
Distance Bandwidth (kms)	20	10	20	10

Notes: This table shows the relationship between being in a command area and Gini and EVI. Standard errors are clustered at district level. There is an increase in Gini coefficient by about 0.90 pp, significant at 5% level within the command area, and is robust to varying bandwidths (10km, 20km) and including full-set of historical covariates $\mathbf{H}_{vds}$, and the running variable (distance to the princely-state border) and its treatment interaction to account for smooth spatial trends across the boundary. Agricultural productivity increases by 10–25% within the command area.

Table A5: Impact of SC Population on Inequality

	Gini, All HHs				
	(1)	(2)	(3)	(4)	(5)
Panel A: Baseline					
Pct Popln SC	12.557***	12.926***	12.420***	12.489***	12.169***
	(1.339)	(1.123)	(0.848)	(0.697)	(0.603)
R-squared	0.029	0.181	0.253	0.343	0.410
N	283272	275037	275037	275037	275000
Panel B: w/Population Control					
Pct Popln SC	11.715***	13.495***	12.278***	11.874***	11.490***
	(1.297)	(1.111)	(0.802)	(0.608)	(0.515)
R-squared	0.126	0.220	0.304	0.394	0.456
N	283181	274946	274946	274946	274909
Agr/History/Market Controls	No	Yes	Yes	Yes	Yes
FEs	No	No	State	District	Subdistrict

Notes: This table gives the impact of SC population share on land inequality for all households. Columns (3)–(5) add state, district, and subdistrict fixed effects. respectively. Columns (2)—(5) include include the full vector of agricultural, market, and historic variables. All specifications include a control for the ST population share. Standard errors are clustered at district level. The effect size indicates that moving from a village at the 25^{th} percentile (SCpopshare$_{vds} = 0.012$) of the SC population to the 75^{th} percentile (SCpopshare$_{vds} = 0.272$) increases the all-household Gini by 4.4% of the mean (71.2).

Table A6: Impact of SC Population on Labor Force

	Ln Popln Density	Ln Male Workers			Pct Male Workers		
		Any Agr	Own-Farm	Ag Labor	Any Agr	Own-Farm	Ag Labor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: State FEs							
Pct Popln SC	0.027	0.006	-0.612***	0.670***	0.004	-0.219***	0.223***
	(0.036)	(0.054)	(0.069)	(0.066)	(0.010)	(0.014)	(0.011)
R-squared	0.427	0.089	0.061	0.195	0.248	0.223	0.272
N	272688	272535	269964	263778	272759	272759	272759
Panel B: District FEs							
Pct Popln SC	0.129***	0.106***	-0.504***	0.733***	-0.011	-0.227***	0.216***
	(0.034)	(0.036)	(0.042)	(0.054)	(0.007)	(0.009)	(0.009)
R-squared	0.487	0.213	0.192	0.276	0.352	0.339	0.351
N	272686	272533	269962	263776	272757	272757	272757

Notes: This table gives the impact of SC population share on demographic characteristics. All regressions control for full set of agriculture, market and historic-related covariates, and district fixed effects. Standard errors are clustered at district level. The outcome variables are log of population density in Column (1), number of male workers employed in agriculture in logs in Column (2), number of male agriculture labourers in Column (4). Columns (5)-(7) uses share of male workers instead of logs. We find that population density is indeed higher when district fixed effects are included, and that this is associated with a larger agricultural wage-labor population.

Table A7: Towns and Inequality

Outcome: Gini Coefficient, All	(1)	(2)	(3)	(4)
Prox Urban	2.943***	1.729***	2.300***	1.288***
	(0.514)	(0.497)	(0.280)	(0.272)
Fitted Agri	2.904***	2.296***	2.372***	0.827*
	(0.664)	(0.578)	(0.477)	(0.453)
Princely	-2.994***	-2.868***	-0.292	-0.065
	(0.994)	(1.036)	(0.370)	(0.361)
Zamindar	1.163	1.460	-0.686	-0.830
	(1.489)	(1.346)	(1.843)	(1.614)
SC	11.197***	11.319***	12.092***	11.298***
	(1.023)	(1.037)	(0.811)	(0.719)
ST	-0.101	0.772	-0.311	1.785**
	(1.354)	(1.474)	(0.904)	(0.858)
Prox Urban X				
Fitted Agri	-1.019***	-0.767**	-0.778***	-0.575***
	(0.354)	(0.317)	(0.215)	(0.195)
Princely	0.408	0.385	0.093	0.176
	(0.659)	(0.635)	(0.350)	(0.321)
Zamindar	0.538	0.737	0.047	0.360
	(0.709)	(0.632)	(0.372)	(0.341)
SC	0.457	0.402	0.682	0.898
	(0.877)	(0.826)	(0.660)	(0.579)
ST	-0.159	0.198	0.318	0.661
	(1.153)	(1.103)	(0.724)	(0.684)
R-squared	0.220	0.280	0.336	0.391
N	260586	260499	260586	260499
Fixed Effects	State		District	
Controls:				
Ln Popln Density	No	Yes	No	Yes

Notes: This table presents the results of a regression of the Gini coefficient (including both landed and landless households) on the indicated variables. *Prox Urban* is an indicator variable taking a value of 1 for villags within 10kms of a town or city; and *Fitted Agri* ($\widehat{\text{EVI}}_{vds}$) is the z-score of the fitted value of NDVI using the agricultural variables, from equation 1a. A control is also included for ln population density. Standard errors are clustered at district level. Columns (1)–(2) include state fixed effects; and column (3)–(4) district fixed effects. We find that the direct impact of fitted agriculture ($\widehat{\text{EVI}}_{vds}$) is mitigated by approximately 1/3 in close proximity to a town. Similar to before, once again all the interaction coefficients with historic factors are insignificant, highlighting the persistence of historic factors in determining inequality.

Table A8: Agricultural Productivity, Market Development, and Inequality

Outcome: Gini Coefficient, All	(1)	(2)	(3)	(4)
Fitted Agri	3.474***	2.901***	2.608***	1.197***
	(0.650)	(0.587)	(0.448)	(0.431)
High Manu	6.466***	5.522***	4.344***	3.446***
	(0.826)	(0.809)	(0.561)	(0.597)
Prox Town	2.248***	1.434***	1.963***	1.354***
	(0.233)	(0.219)	(0.157)	(0.150)
Prox Road	1.790***	1.577***	1.237***	0.988***
	(0.290)	(0.284)	(0.180)	(0.174)
Prox RR Station	2.004***	1.308***	1.712***	1.161***
	(0.220)	(0.212)	(0.165)	(0.157)
Bank Present	9.049***	6.197***	8.764***	5.925***
	(0.330)	(0.309)	(0.281)	(0.237)
Mandi Present	4.624***	3.379***	4.028***	2.741***
	(0.404)	(0.366)	(0.309)	(0.255)
Fitted Agri X				
High Manu	-1.470*	-1.928**	-1.384**	-1.686**
	(0.783)	(0.776)	(0.687)	(0.762)
Town	-0.890***	-0.676**	-0.765***	-0.593***
	(0.283)	(0.276)	(0.194)	(0.186)
Road	0.130	0.134	-0.034	0.018
	(0.304)	(0.315)	(0.182)	(0.176)
RR Station	-0.036	0.049	-0.111	-0.063
	(0.246)	(0.246)	(0.199)	(0.194)
Bank	-0.376	-0.287	-0.197	-0.124
	(0.348)	(0.308)	(0.299)	(0.234)
Mandi	-1.919***	-1.642***	-1.780***	-1.322***
	(0.361)	(0.297)	(0.350)	(0.246)
R-squared	0.255	0.300	0.357	0.401
N	260450	260450	260450	260450
Fixed Effects	State		District	
Controls:				
Ln Popln Density	No	Yes	No	Yes
Town Prox w/Interactions	Yes	Yes	Yes	Yes

Notes: This table presents the results of a regression of the all household Gini coefficient on the indicated variables. *High Manu* is a dummy (=1), indicating high share of the non-agriculture working population. *Prox Road*, *Prox Town*, *Prox RR Station*, are dummies (=1) for villages falling within 5 kms, 10km, and 5km of a major highway, town/city, and railroad, respectively. *Bank Present* and *Mandi* are dummies (=1) if the village has any bank or agricultural market. *Fitted Agri* ($\widehat{\text{EVI}}_{vds}$) is the z-score of the fitted value of NDVI using the agricultural variables, from equation 1a, and is interacted with all the variables. Standard errors are clustered at district level. A control is also included for ln population density. We see that non-agricultural labor force and town proximity have independent impacts on the influence of agricultural productivity; and that the presence of a *mandi* also reduces the influence of agricultural productivity. However, we find no evidence that road or railway station proximity, or bank presence, has any impact on the influence of agricultural productivity.

Table A9: Inequality and Public Goods

	Gini All HHs		Gini Landowning HHs			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Govt Primary School						
Gini	0.078***	0.732***	0.181***	0.593***	0.182***	0.594***
	(0.011)	(0.065)	(0.008)	(0.029)	(0.008)	(0.029)
Giniˆ2		-0.504***		-0.430***		-0.430***
		(0.048)		(0.029)		(0.029)
Pct HHs Landless			-0.004	0.009	0.002	0.014*
			(0.007)	(0.007)	(0.007)	(0.007)
R-squared	0.294	0.297	0.298	0.301	0.298	0.301
N	265295	265295	265305	265305	265252	265252
Panel B: Paved Road						
Gini	0.111***	0.204***	0.152***	0.282***	0.149***	0.288***
	(0.013)	(0.061)	(0.011)	(0.032)	(0.011)	(0.032)
Giniˆ2		-0.072		-0.136***		-0.146***
		(0.043)		(0.029)		(0.029)
Pct HHs Landless			0.049***	0.053***	0.032***	0.036***
			(0.009)	(0.009)	(0.009)	(0.009)
R-squared	0.313	0.313	0.314	0.314	0.315	0.315
N	264970	264970	264980	264980	264929	264929
Panel C: Primary Health Center						
Gini	0.084***	-0.111***	0.069***	0.083***	0.066***	0.088***
	(0.006)	(0.019)	(0.004)	(0.011)	(0.004)	(0.011)
Giniˆ2		0.150***		-0.015		-0.022**
		(0.014)		(0.011)		(0.011)
Pct HHs Landless			0.061***	0.062***	0.047***	0.048***
			(0.004)	(0.004)	(0.003)	(0.003)
R-squared	0.053	0.054	0.055	0.055	0.060	0.060
N	264763	264763	264773	264773	264721	264721
Panel D: Sanitation Campaign						
Gini	0.047***	-0.015	0.038***	0.085***	0.036***	0.087***
	(0.010)	(0.035)	(0.008)	(0.019)	(0.007)	(0.019)
Giniˆ2		0.048*		-0.050***		-0.053***
		(0.026)		(0.018)		(0.018)
Pct HHs Landless			0.029***	0.030***	0.022***	0.024***
			(0.006)	(0.006)	(0.006)	(0.006)
R-squared	0.438	0.438	0.438	0.438	0.439	0.439
N	264802	264802	264812	264812	264760	264760
Control Non-Agr Labor					Yes	Yes

Notes: This table presents the relationship between four types of public goods – presence of government primary goods, paved roads, primary health centers and sanitation campaign– in Panels A-D, respectively and land inequality. The $\hat{\beta}_{i1}$ and $\hat{\beta}_{i2}$ coefficients from equation 7 are plotted. We control for our full vector of agriculture, market, and historical-related variables, population density, and district fixed effects. Consistent with the literature (Banerjee and Somanathan, 2007), we also include social fractionalization index. Standard errors are clustered at the district level. Columns (1)–(2) take as the explanatory variable the Gini for all households; and columns (3)–(6) the Gini for landowning households. We find a non-linear relationship – public goods are positively associated with the land inequality with a declining impact negative coefficients on Gini-square.

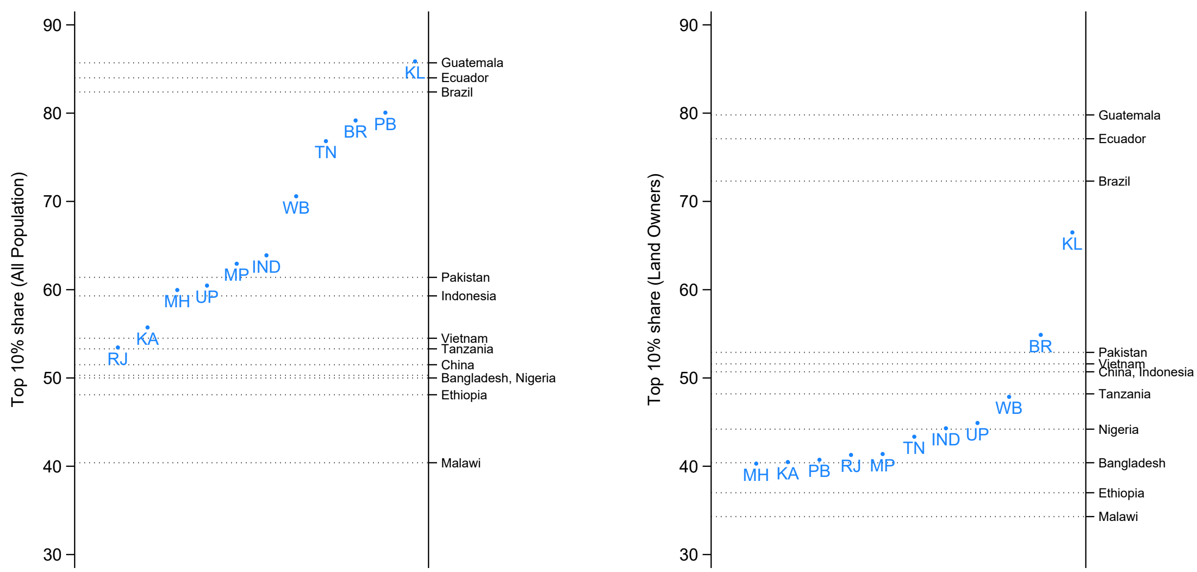
Table A10: Princely State Border Discontinuity Design: Excluding Outliers

	Land Gini, All Households					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
Princely State	-2.211**	-2.282**	-2.526***	-2.616***	-2.320***	-2.044***
	(1.064)	(0.999)	(0.810)	(0.756)	(0.707)	(0.784)
R-squared	0.227	0.225	0.278	0.284	0.380	0.400
N	43541	21799	43541	21799	43520	21782
Panel B: Removing above p95						
Princely State	-3.055**	-2.998**	-2.990***	-2.865***	-2.552***	-2.133***
	(0.929)	(0.916)	(0.697)	(0.731)	(0.603)	(0.809)
R-squared	0.180	0.181	0.230	0.237	0.327	0.351
N	41036	20554	41036	20554	41012	20530
Panel C: Removing above p90						
Princely State	-3.402*	-3.162**	-3.204**	-2.845***	-2.665***	-2.236***
	(0.905)	(0.922)	(0.688)	(0.759)	(0.629)	(0.859)
R-squared	0.162	0.164	0.214	0.222	0.308	0.335
N	39191	19655	39191	19655	39168	19635
Panel C: Removing above p75						
Princely State	-3.179*	-3.089**	-2.828**	-2.594***	-2.324***	-2.197***
	(0.805)	(0.896)	(0.714)	(0.809)	(0.708)	(0.946)
R-squared	0.125	0.127	0.174	0.184	0.262	0.292
N	33602	16919	33602	16917	33581	16898
BW (kms)	20	10	20	10	20	10
State FEs	Yes	Yes	Yes	Yes	Yes	Yes
Princely State FEs			Yes	Yes		
Segment FEs					Yes	Yes

Notes: This table shows the results of a regression of all-household Gini on an indicator variables for villages located within the princely state. Panel A excludes control variables; Panel B includes the vector of agricultural variables ($\mathbf{A}_{vds}$); Panel C includes both agricultural and historic (caste) variables; and Panel D further adds the running variable (distance to the princely-state border) and its treatment interaction to account for smooth spatial trends across the boundary. Standard errors are clustered at district level. The coefficients are relatively stable across specifications with princely states having approximately 2–3 percentage points lower land inequality.

Appendix B

Figure B1: Land Area Inequality in Comparison with Other Countries

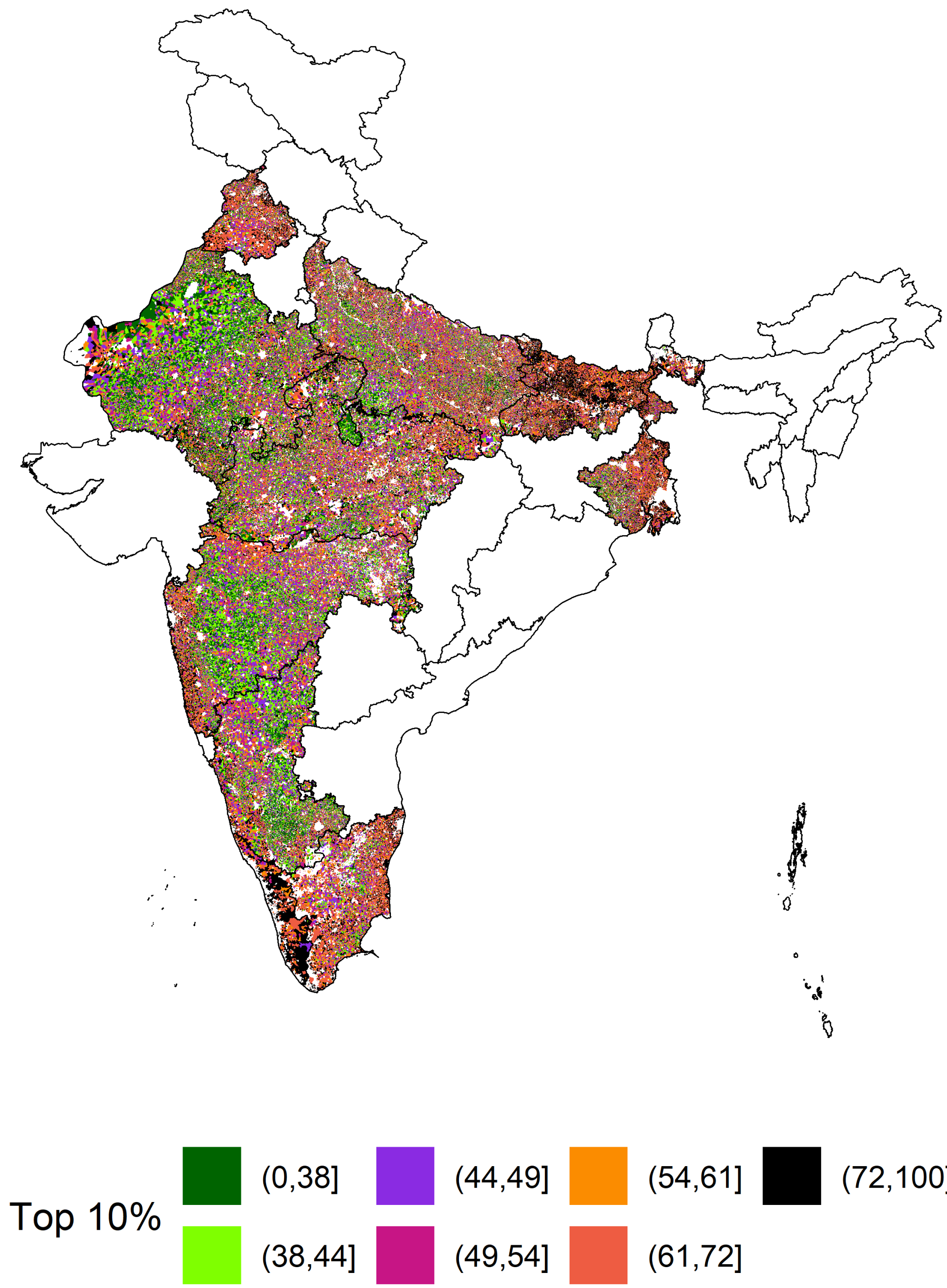


(a) Top 10% shares (All population)

(b) Top 10% shares (within landowners)

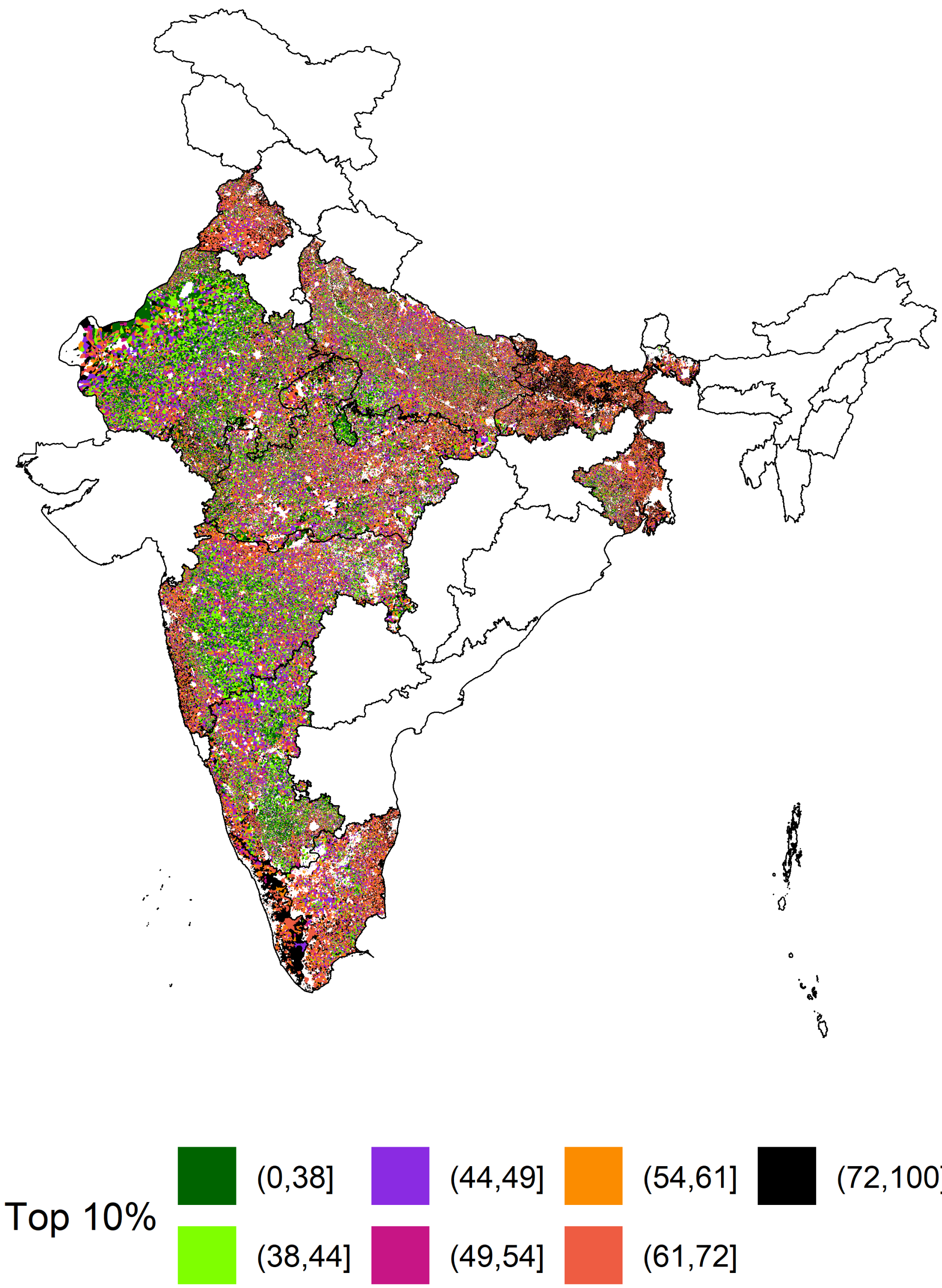
Notes: The figure represents the average village-level land inequality (top 10% shares) among the total population and among landowners (excluding the landless) in ten large Indian states, compared with other countries (Bauluz et al., 2025). RJ—Rajasthan; KA—Karnataka; UP—Uttar Pradesh; MH—Maharashtra; MP—Madhya Pradesh; WB—West Bengal; TN—Tamil Nadu; BR—Bihar; PB—Punjab; KL—Kerala.

Figure B2: Land inequality: Top 10% shares, all households



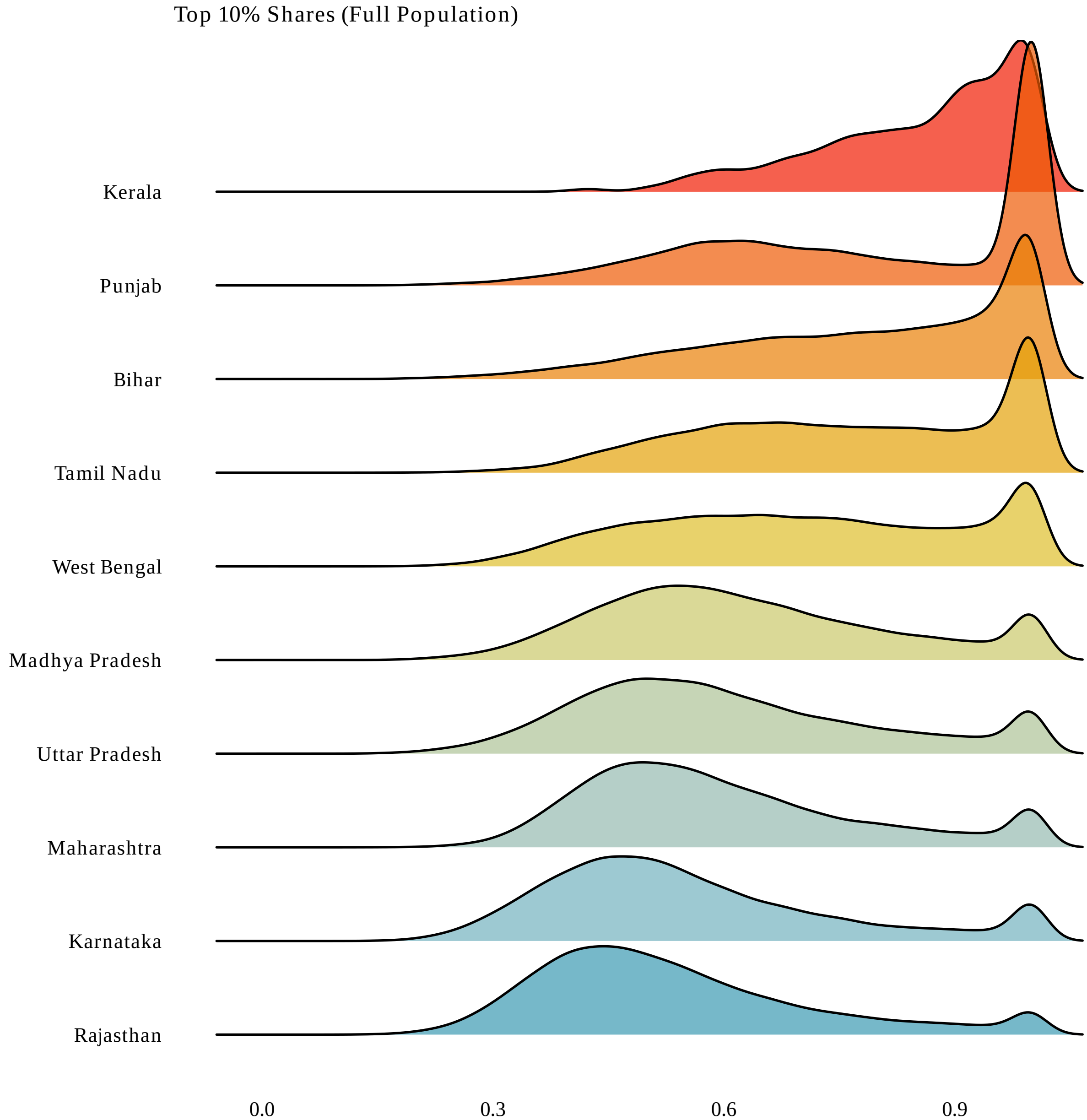
Notes: The figure plots the distribution of agricultural land: top 10% shares in full population in Indian villages.

Figure B3: Land inequality: Top 10% shares, landowners



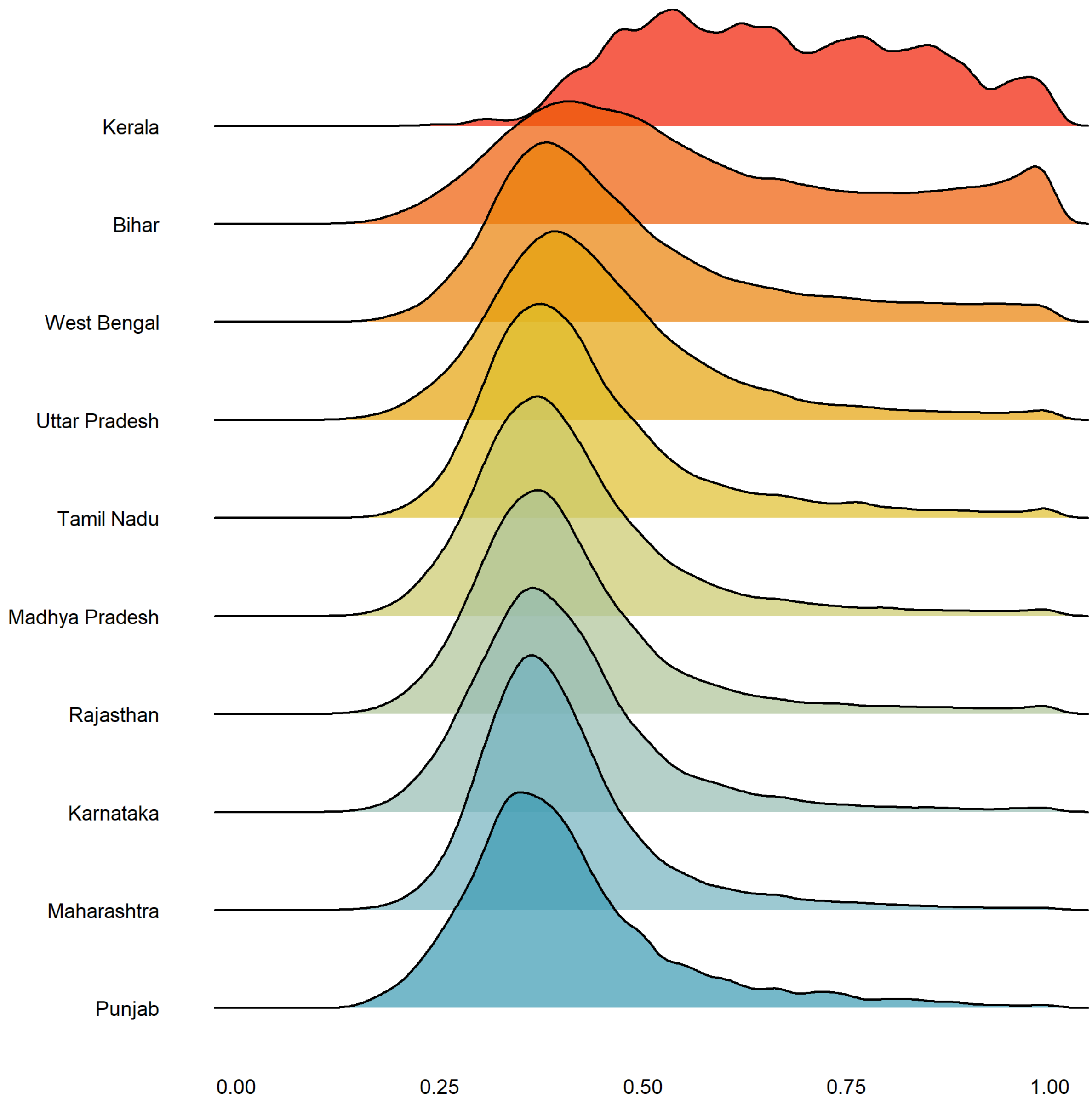
Notes: The figure plots the distribution of agricultural land: top 10% shares in landowners in Indian villages.

Figure B4: **Top 10% Shares, All households**



Notes: The figure plots the distribution of agricultural land shares for the top 10% all households in different states of India.

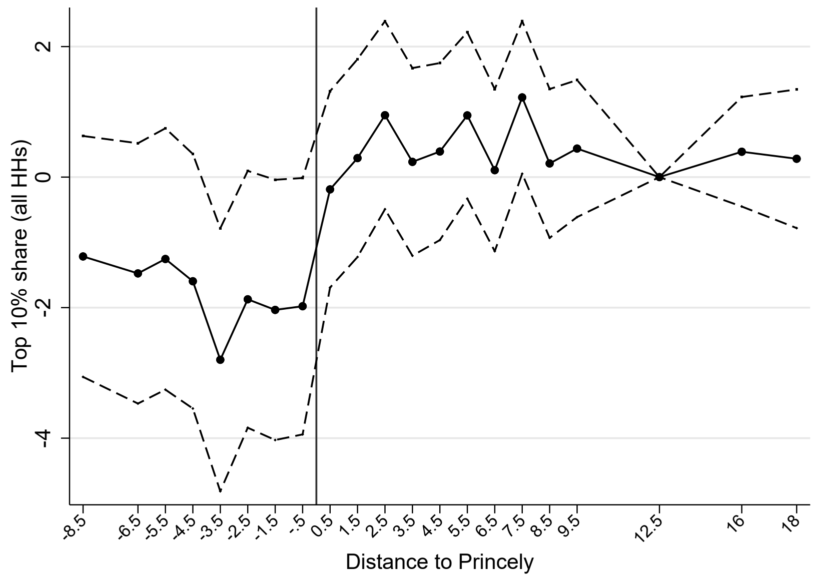
Figure B5: **Top 10% Shares, landowning households**



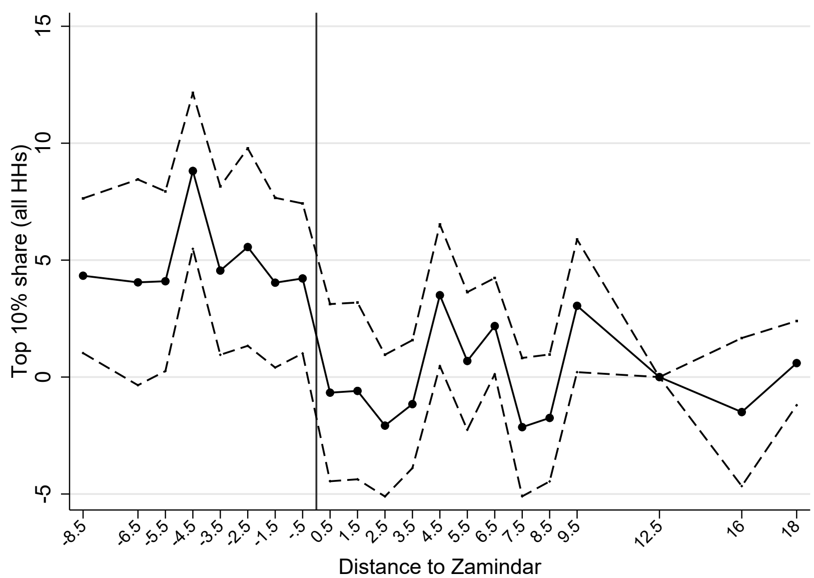
Notes: The figure plots the distribution of agricultural land shares for the top 10% landowners households in different states of India.

Figure B6: Distance to Institutions: Top 10% (all HH's)

(a) Princely



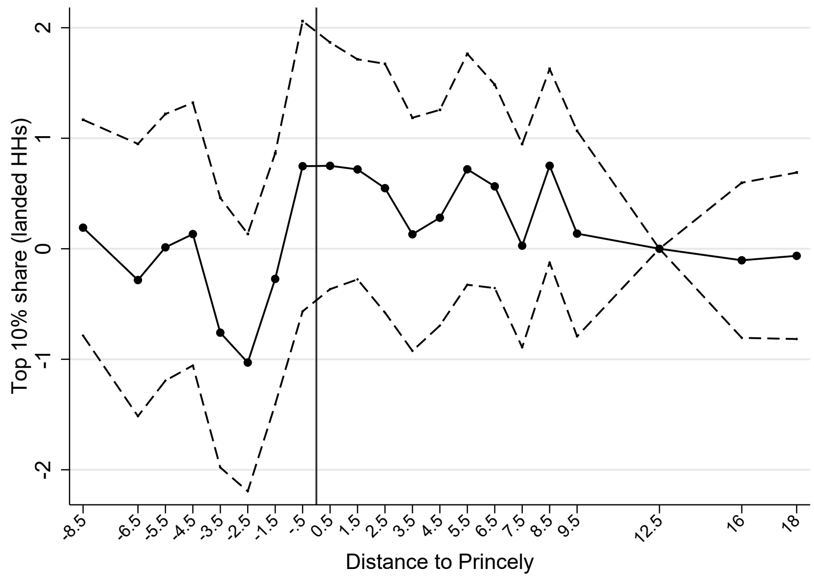
(b) *Zamindar*



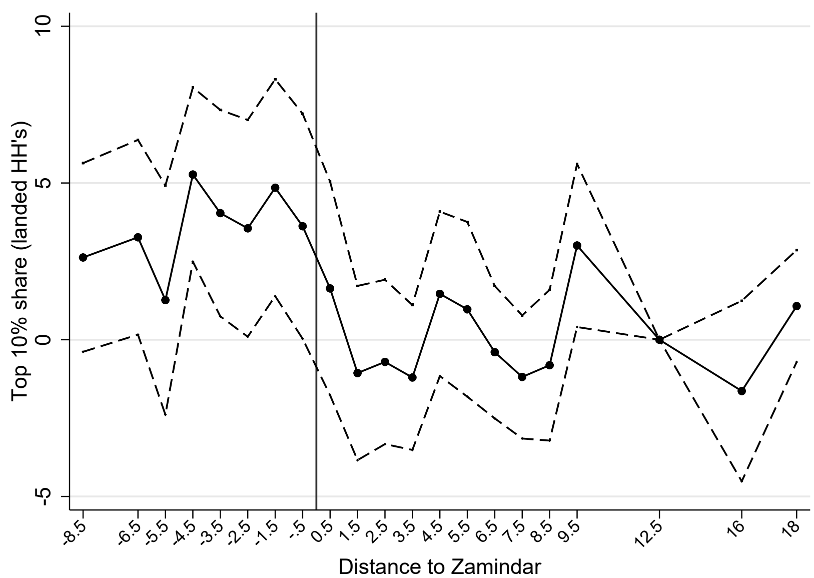
Note: Figure B6 plots the relationship between distance to princely states and *zamindar* areas and top 10% land shares **among full population** in top and bottom graphs respectively. The coefficients are following equation 6, where we control for vectors of agricultural ($\mathbf{A}_{vds}$), as well as 20-km boundary segment fixed effects. For princely state analysis we additionally include state fixed effects. For *zamindar* analysis we exclude state fixed effects and include the market variables ($\mathbf{M}_{vds}$), caste variables, and supplemental control variables. Standard errors are clustered at the boundary segment level. Negative distance denotes **inside** princely states or *zamindari* area and positive values denote **outside**. Princely states are associated with decreases in top 10% shares, and *zamindar* areas with increases.

Figure B7: Distance to Institutions: Top 10% (landed HH's)

(a) Princely



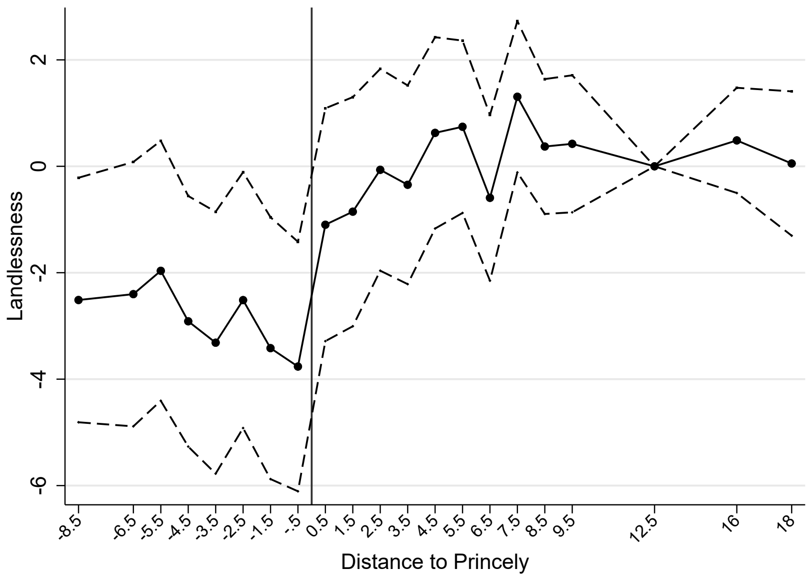
(b) *Zamindar*



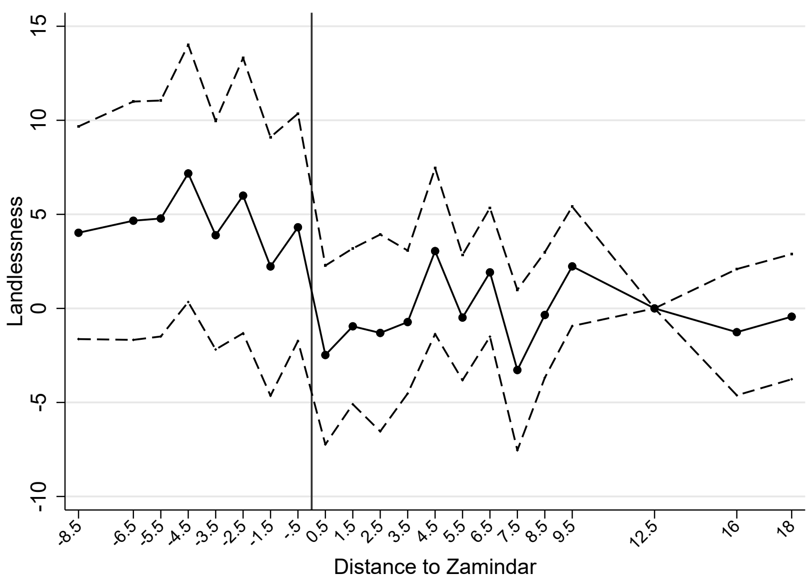
Note: Figure B7 plots the relationship between distance to princely states and *zamindar* areas and top 10% land shares **within landed households** in top and bottom graphs respectively. The coefficients are following equation 6, where we control for vectors of agricultural ($\mathbf{A}_{vds}$), as well as 20-km boundary segment fixed effects. For princely state analysis we additionally include state fixed effects. For *zamindar* analysis we exclude state fixed effects and include the market variables ($\mathbf{M}_{vds}$), caste variables, and supplemental control variables. Standard errors are clustered at the boundary segment level. Negative distance denotes **inside** princely states or *zamindari* area and positive values denote **outside**. Princely states are associated with decreases in top 10% shares, and *zamindar* areas with increases.

Figure B8: Distance to Institutions: Landless shares

(a) Princely

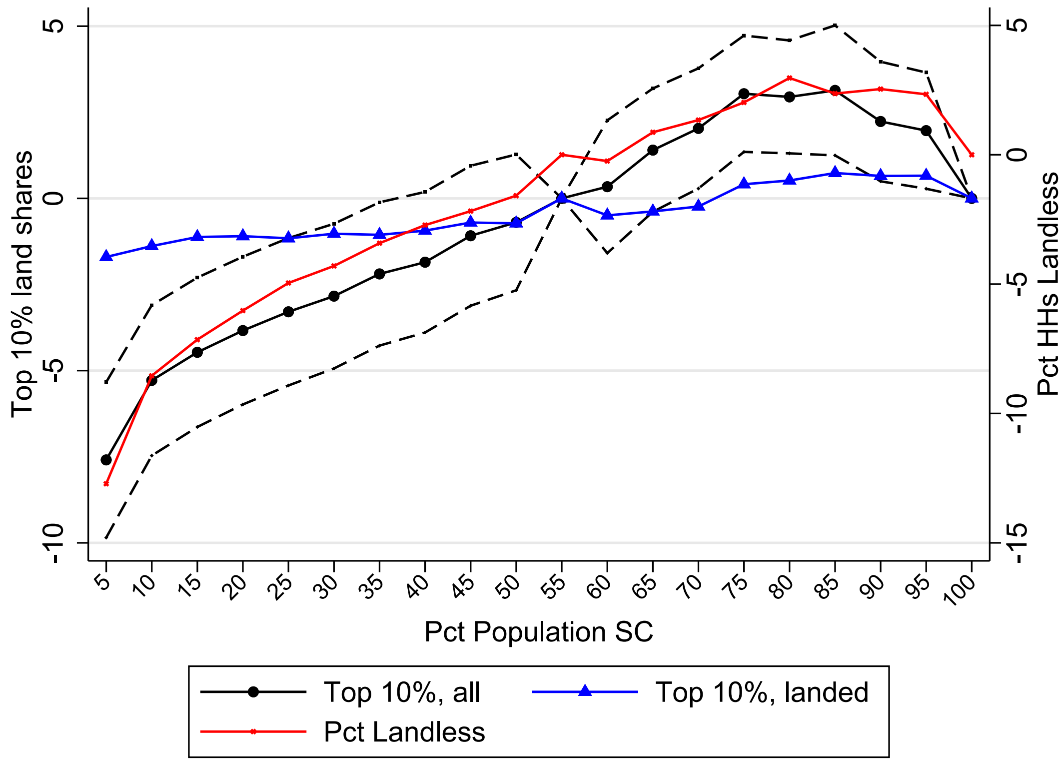


(b) *Zamindar*



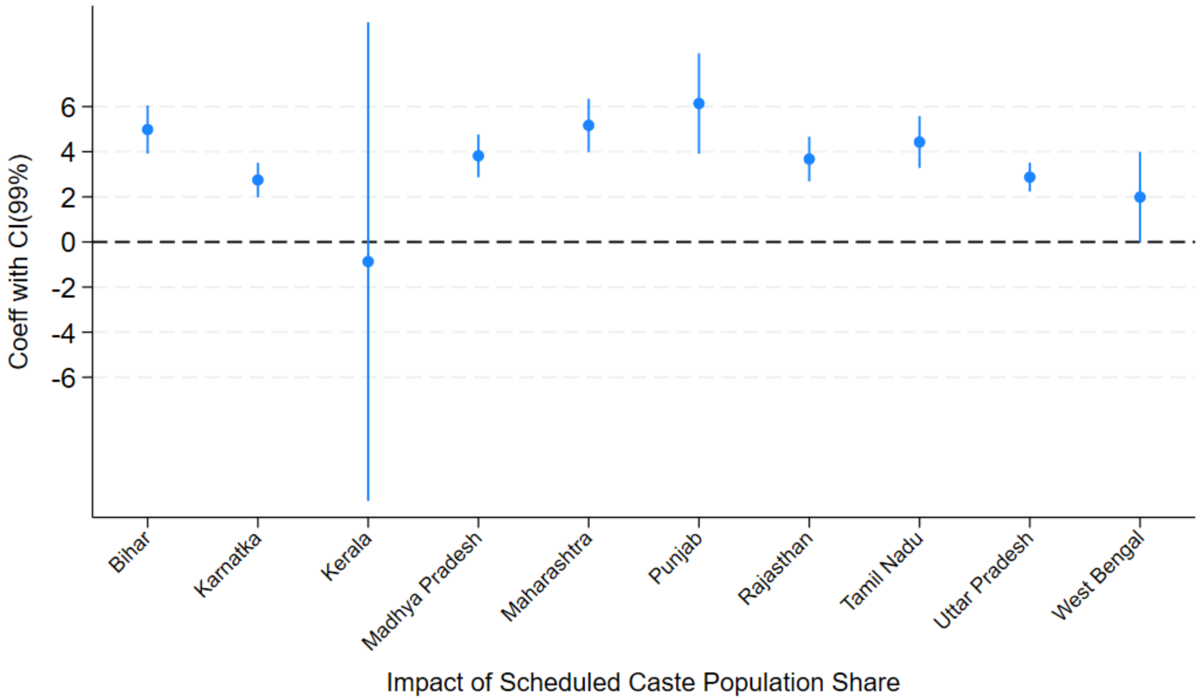
Note: Figure B8 plots the relationship between distance to princely states and *zamindar* areas and **landless households** in top and bottom graphs respectively. The coefficients are following equation 6, where we control for vectors of agricultural ($\mathbf{A}_{vds}$), as well as 20-km boundary segment fixed effects. For princely state analysis we additionally include state fixed effects. For *zamindar* analysis we exclude state fixed effects and include the market variables ($\mathbf{M}_{vds}$), caste variables, and supplemental control variables. Standard errors are clustered at the boundary segment level. Negative distance denotes **inside** princely states or *zamindari* area and positive values denote **outside**. Princely states are associated with decreases in top 10% shares, and *zamindar* areas with increases.

Figure B9: **Caste and Inequality**



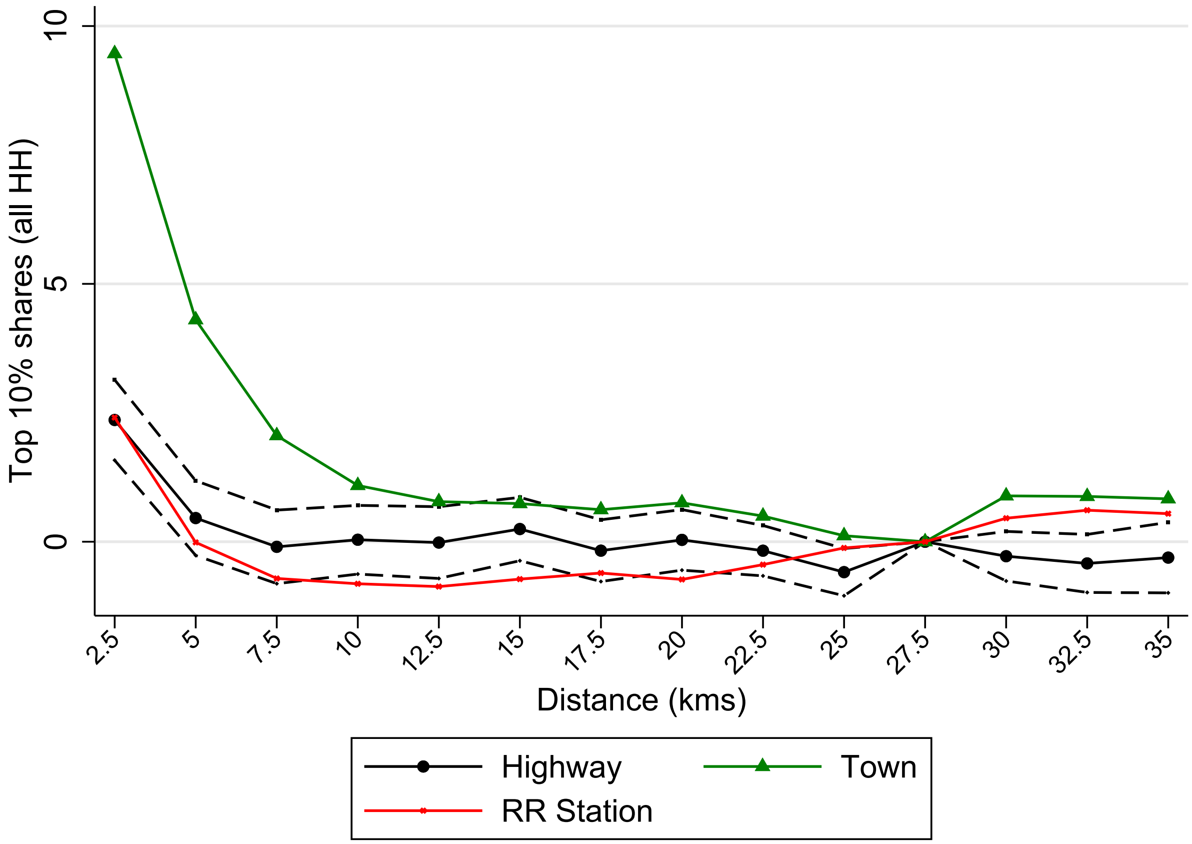
Note: Figure B9 plots the relationship between percent of SC population in a village and three land inequality measures – top 10% shares in full population, top 10% shares within landed population, and share of landless households. Standard errors are clustered at district level. The relationship for all-household top 10% shares is positive and largely linear in the share of the population being SC between 5–80%; is far steeper when the share increases from 5–10%; and declines for values greater than 80%. This relationship is largely driven by landlessness, with the relationship between the landed-household top 10% shares and the SC population, is somewhat negative. The plotted coefficients ($\hat{\beta}_i$) are following equation 5, where we control full set of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$)- related variables, along with total village land area, and district fixed effects.

Figure B10: **Impacts of SC Population Share on Top 10% shares (all households)**



Note: Figure B10 plots the relationship between the Top 10% shares (all households) and standardized Scheduled Caste population share (by state's share and standard deviation), separately by state. The coefficients for regressions run separately for individual states. District fixed effects are included, as well as the agricultural and demographic variables, and (ln) population density. Standard errors are clustered at district level. For most states, changing the SC population by 1 standard deviation, increases the top 10% shares between 4–6 percentage points and is statistically significant at 1% level. One notable exception to this pattern is the states of Kerala and West Bengal, which were long governed by left-wing parties, and in which post-independence land reforms are generally considered to have been the most successful: in these states, the relationship between SC population share and land inequality is substantially smaller and is statistically insignificant.

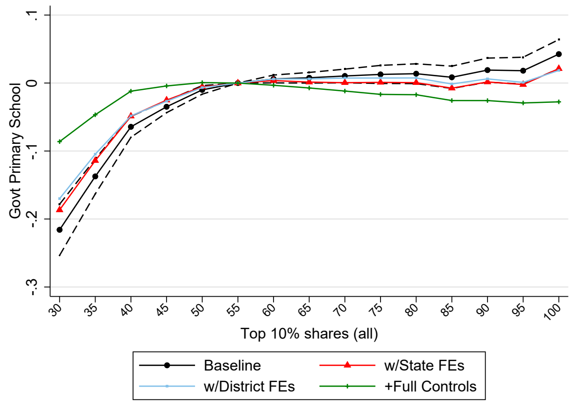
Figure B11: **Markets and Inequality**



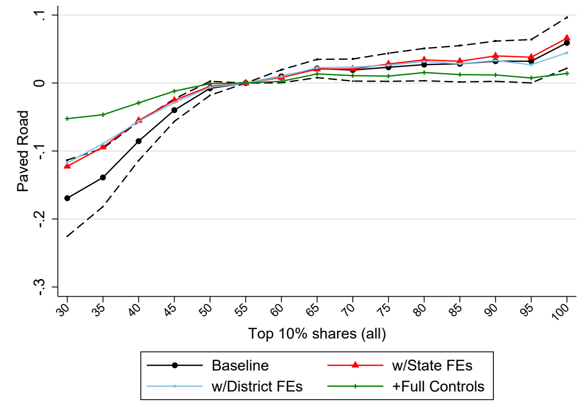
Note: Figure B11 plots the relationship between the top 10% shares (all households) and distance to: major highways, and railroad stations. For each market-related variables, the Gini coefficient increases substantially as proximity to the respective locations increases. Standard errors are clustered at district level. The increase is the largest for towns. For roads and railway stations the relationship is smaller, but still positive. Moreover, we see that inequality is elevated up to further distances from towns (10kms) than roads and railway stations (2.5kms). The plotted coefficients ($\hat{\beta}_i$) are following equation 6, where we control full set of agricultural ($\mathbf{A}_{vds}$), market ($\mathbf{M}_{vds}$), and historic ($\mathbf{H}_{vds}$)- related variables, along with total village land area, and district fixed effects.

Figure B12: Inequality (top 10% shares) and Public Goods

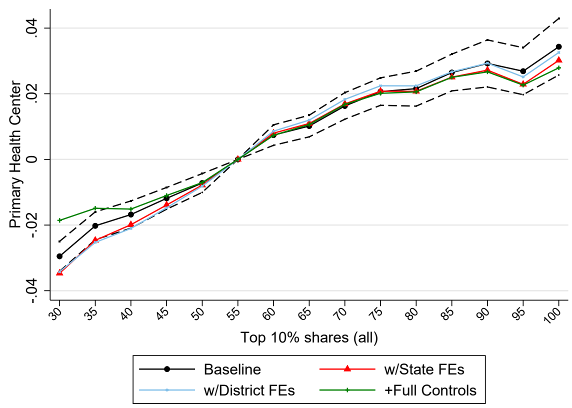
(a) Govt Primary School



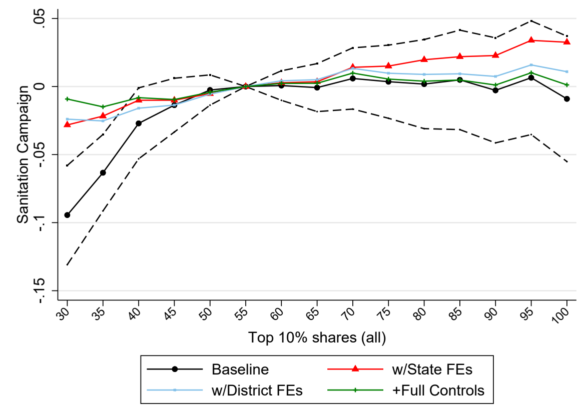
(b) Paved Road



(c) Primary Health Center



(d) Sanitation Campaign



Note: Figure B12 plots the relationship between top 10% shares and the presence of the indicated public goods. Results are shown with and without control variables. The full controls include the agriculture, market, and history variables, as well as ln population density and social fractionalization index (following Banerjee and Somanathan, 2007). Standard errors are clustered at district level. While the raw correlation is generally upwards sloping and concave, the inclusion of controls leads to an inverted-U relationship for some outcomes. This suggests that intermediate levels of inequality may promote the provision of public goods, but that extreme inequality is detrimental to public goods provision.