Does Market Integration Increase Rural Land Inequality? Evidence from India

Claudia Berg. The World Bank[1]

Brian Blankespoor, The World Bank

M. Shahe Emran, IPD, Columbia University

Forhad Shilpi, DECRG, World Bank

First Circulated Draft: July 22, 2022
This Version: April 26, 2023

ABSTRACT

Do rural-urban highways increase land inequality in villages? Theory suggests that, with credit market imperfections, lower trade costs can increase land inequality through increasing-returns technology adoption. Using data on household land ownership in rural India, we provide the first evidence on this issue. Identification exploits the distance of a district to the Golden Quadrilateral network (inconsequential place) and the length of colonial railroad in the 1880s in a district. A 10 percent increase in market access of a district increases land Gini by 2.5 percent, share of landless households by 6.8 percent, and adoption of increasing-returns farming technology by 3.5 percent.

JEL Codes: O18, O12, D30, D63, Q15, R40

Key Words: Market integration, Transport Infrastructure, Land inequality, Landlessness, Gravity Measure, Rural India, Colonial Railroad, Golden Quadrilateral, Credit Market Imperfections, Increasing Returns Farming Technology

[1] We would like to thank Treb Allen for helpful discussions on market access data, Lakshmi Iyer for help with understanding the colonial land revenue system in India, and Latika Chaudhury for population density data at the district level. We also thank Simon Alder, Dilip Mookherjee, Steven (Ruoyu) Chen, Harris Selod, Saher Asad, Raghbendra Jha, Hrishikesh Prakash Patel for discussions and/or comments. Standard disclaimers apply. Email for correspondence: shahe.emran.econ@gmail.com.

(1) Introduction

Public investment in transport and communications infrastructure experienced a sustained increase in developing countries in recent decades, as many governments identified spatial isolation from urban markets as a primary factor behind high poverty rates in lagging hinterlands and widening regional inequality (Kanbur and Venables (2005), World Bank (2005)). Donors and multilateral development banks such as World Bank and Asian Development Bank allocated a substantial proportion of their aid and loan portfolio to transport infrastructure projects in developing countries. Using data from the AidData, Tierney et al. (2011) estimate that, from 1948-2013, over $701 billion were committed for transport and storage sector across the world of which about 7% was in India.[2] The huge investments in roads and highways, bridges, and railways lowered the transport costs substantially and led to spatial market integration. According to a gravity measure, average market access in India increased by 25 percent each decade from 1962 to 2011.[3]

Does integration of a village with urban markets unleash economic forces that lead to higher land inequality? The existing theoretical analysis suggests that lower trade costs can lead to a concentration in land ownership because of credit market imperfections. Braverman and Stiglitz (1989) develop a model where lower trade costs increase returns to land by increasing producer prices irrespective of farm size, but also improve profitability of technology adoption. Since large farmers have better access to credit to finance technology adoption, they buy land from credit-constrained small farmers, especially when the technology involves increasing returns such as irrigation and mechanization of cultivation and harvesting.[4]

Many countries attempted land reform in the 1950s-1970s, but recent evidence suggests that land inequality has increased in most of the countries after economic liberalization in the 1980s (Bauluz et al. (2020)).[5] The effects of market liberalization that swept the developing world in the 1980s and 1990s depend on the domestic trade costs which are largely determined

[2]From 1995- 2014, World Bank *disbursed* over $94 Billion in the transport and storage sector to developing countries of which nearly more than 10% (nearly $10 Billion) was in India (Aid Data, 2017).

[3]This is based on the estimates of market access in India provided by Treb Allen.

[4]Large land owners may also get a lower interest rate implying higher capitalization of the land rents. This would make the land more valuable to large land owners even without technology adoption.

[5]Ghatak and Roy (2007) present evidence that many decades of land reform policies in India failed to reduce land inequality significantly.

by the location of a village relative to the urban centers and the availability and quality of transport infrastructure.[6] A highly skewed land ownership distribution in a country not only reflects inequality, but can adversely affect economic efficiency, public goods provision, entrepreneurial development, and political stability (Binswanger et al. (1995), Sokoloff and Engerman (2000), Mariscal and Sokoloff (2000), and Banerjee and Newman (1993)).

Using household-level data on land ownership in rural districts from the India Human Development Survey II in 2012 (henceforth IHDS-II), we provide the first (to our knowledge) empirical analysis of the effects of market integration on land inequality, and explore the role of technology adoption as a mechanism.[7] India is an excellent case study to understand the effects of trade costs on land inequality. India is a vast country with a long history of transportation network and substantial geographic variation: an index of market access in 1996 at the district level varies from 3.5 to 100. This spatial variation in market access helps to identify the effects of market integration. The effects of transport infrastructure on land inequality are also important from a policy perspective as land inequality and land reform have been central policy issues in India for many decades (Bardhan et al. (2014), Ghatak and Roy (2007)).

The empirical analysis uses two measures of land inequality in 2012: a district level land Gini for household land ownership and the proportion of landless households in a district. As a measure of adoption of increasing returns technology in agriculture, we use the proportion of households in a district owning at least one of the following types of farming equipment: tube well, electric pump, diesel pump, tractor, and thresher. We also analyze the effects of market integration on land sales, but the empirical results are not robust. For market integration, we calculate a gravity measure using travel time through roads and highways. Our main results are based on travel time in 1996, but we report results using travel time in 2004 in

[6]For evidence that the transmission of price signals in developing countries depend on the remoteness of a location, see the analysis by Atkin and Donaldson (2015) in the context of Ethiopia and Nigeria. Market liberalization means little for a village which remains in autarkic equilibrium because of a lack of transport infrastructure.

[7]We rely on IHDS-II because it is a high quality household survey. It is not possible to use agricultural census data because of unavailability of land ownership information. See Bauluz et al. (2020) for a discussion on why agricultural census data are not suitable and the advantages of household surveys in studying *land ownership* inequality.

the online appendix.[8] It is important to note that our identification approaches are designed to understand the long-term effects. If the identifying variations are successful in capturing long-term changes, then the estimates should not vary substantially whether we use the 1996 or 2004 market access measure for our analysis.[9]

We develop an instrumental variables approach for potential endogeneity of market access by exploiting two sources of exogeneous variation: colonial railroads in the 1880s and Golden Quadrilateral (GQ) highway network. Motivated by Duranton and Turner (2012) and Donaldson (2018), the location of historical railroad infrastructure in the 1880s is used for estimating the effects of better market access via roads and highways on land inequality in 2012. We discuss (and deal with where it is necessary) potential threats to the exclusion restriction imposed on historical railroad: (i) geographic targeting by the colonial government for poverty alleviation and tax revenue, (ii) commercial motives for the railroads financed by private British investors, and (iii) potential long-term effects of colonial railroads through agglomeration and persistence. We also check whether colonial railroad captures partly the effects of colonial land revenue systems on land inequality (Banerjee and Iyer (2005)). For details, please see section (4.1) below. As a second source of identification, we exploit the distance of a district from the relevant arm of the Golden Quadrilateral highway network (henceforth GQ) in an inconsequential place design. To strengthen the credibility of the research design, we implement three steps: (i) exclude the districts located in the main nodes connected by the GQ, (ii) construct two hypothetical GQ networks using Euclidean distance and least cost path algorithm, (iii) construct hypothetical feeder roads connecting a district center to the nearest GQ arm, again using Euclidean distance and least cost path algorithm. Our identifying instrument is based on double hypothetical routes: the *hypothetical* distance from the district center to the nearest arm of the *hypothetical* GQ network. For details, please see section (4.2) below. To check whether the main conclusions are robust to local violation of the strict exclusion restriction imposed on an instrument (i.e., when there is a small direct

[8]The years 1996 and 2004 are chosen because market access estimates are available for these years from Allen and Atkin (2016). We are grateful to Treb Allen for sharing the market access data and helpful explanations.

[9]It is well-known that, significant changes in land inequality in India may take many decades. This is borne out by a comparison of the land Gini estimates from 2005 and 2012 rounds of IHDS: there is virtually no change in the district level land Gini over the 7-year period. Details are available from the authors.

impact of an instrument on the outcome variable), we implement the "plausibly exogenous" approach of Conley et al. (2012).

The empirical results from 2SLS, Lasso-IV show that market integration increases land Gini and the proportion of landless in a district. Estimates from Lasso-IV suggest that a 10 percent higher market access increases land Gini by 2.55 percent, and the proportion of landless by 6.78 percent in a district (both estimates are significant at the 1 percent level). Evidence on modern (increasing returns) technology adoption provides support for the Braverman-Stiglitz (1989) hypothesis: a 10 percent higher market access increases the adoption of increasing-returns farming technology by 3.5 percent (significant at the 1 percent level).[10] The estimated effect of higher market integration on the incidence of land sales is positive and substantial in magnitude, but it lacks precision in the IV regressions (not significant at the 10 percent level).[11] These conclusions are robust to the relaxation of the exact exclusion restrictions on the instruments imposed in the IV estimation, using the Conley et al. (2012) approach, and the use of alternative values of the trade elasticity parameter (1.5, 3.8), following Allen and Atkin (2016), and travel time year (1996, 2004) for calculating market access.

We also find that the main conclusions are robust to an alternative empirical approach that does not rely on any exclusion restrictions. In particular, we implement Oster (2019) bias adjusted OLS which extends the Altonji et al. (2005) approach of exploiting selection on observables as a guide to selection on unobservables.[12] We also find that the estimated effects of market integration are not driven by spatial heterogeneity in demographic pressure, differences in colonial land revenue systems across districts, differences in land reform policies across states, or differences in land inheritance laws and customs between Hindu and Muslim

[10]Adoption of increasing returns technology is expected to increase land productivity. However, note that the large landowners are not necessarily the best farmers (in terms of farming knowledge accumulated over generations, for example). If some of the relatively inefficient large farmers drive out more efficient small farmers because of credit market imperfections, this would lead to misallocation and productivity loss compared to a benchmark with equal access to credit for small and large farmers.

[11]The OLS estimate with state fixed effects suggests that a 10 percent higher market access increases the incidence of land sales by 8 percent (significant at the 5 percent level).

[12]For recent applications of the Oster (2019) approach to estimating causal effects, see, for example, van Maarseveen (2020). However, the Oster's bias adjusted OLS estimates should be treated as lower bounds in our application as this approach does not correct for attenuation bias owing to measurement error in the market access variable.

population.[13]

The rest of the paper is organized as follows. Section (2) provides a discussion on the country background, focusing on economic and land inequality and the development of transport infrastructure. We discuss the related literature in section (3) with a focus on India. Section (4) develops the empirical strategy for identifying the effects of market integration on land inequality. Section (5) provides a discussion on the household survey data we use from the Indian Human Development survey and the construction of the main variables including the gravity measure of market access. The next section (6) is devoted to the estimation results from our empirical analysis. Section (7) offers the evidence on the mechanisms and tests whether the Braverman-Stiglitz hypothesis is rejected by data. The paper ends with a set of conclusions summarizing the main findings and the contributions of the paper to the existing literature.

(2) Country Background

Inequality in India

A substantial body of evidence suggests that income and wealth inequality increased in India in recent decades. According to the estimates reported by Himanshu (2019), consumption Gini increased from 0.30 in 1983 to 0.37 in 2011-2012 (based on NSS data). Evidence suggest that wealth inequality also increased: the share of top 10 percent grew from 45 percent in 1981 to 65 percent in 2012. The most important component in the wealth portfolio of Indian households is land (farming land and house). Over the years, land contributed more than 60-65 percent of the total household wealth; land and building combined forming around 85-90 percent of the total household wealth (Bharti (2018)).

After independence, India adopted a socialist economic system, nationalizing the industrial sector, and imposing restrictions on international trade. But the vast swath of the agricultural economy was never seriously considered for public ownership (unlike China), the distributional concerns were to be addressed by land and tenancy reform. Over the decades, various tenancy reform and redistributive land reforms imposing ceilings were implemented, and the policies

[13] Bardhan et al. (2014) provide evidence that demographic pressure and inheritance play important roles in shaping land inequality in the state of West Bengal in India. Banerjee and Iyer (2005) find that colonial land revenue system had long-term effects on land inequality, agricultural productivity and irrigation in a district.

vary widely across different states (Besley and Burgess (2000)). We control for these state level variation in land policies by including state fixed effects. There is substantial evidence that the land reform policies failed to reduce land inequality. Even in the state of West Bengal which implemented perhaps the most comprehensive tenancy reform, evidence suggests that land inequality did not decline after the implementation of the reforms, the forces of demographic pressure and land inheritance law overwhelming the effects of land reform (Bardhan et al. (2014)). Ghatak and Roy (2007) report evidence that land reform in India was not effective in reducing land inequality. Besley et al. (2016) reach somewhat different conclusions: land inequality is lower in areas that saw greater intensity of tenancy reform over 3 decades with heterogeneity across caste groups.

Transport Development in India

The transport sector in India has experienced dramatic growth in the post independence period with important changes in the mode of transport for both freight and passenger traffic. The freight transport volume by roads and highways increased from 12.09 billion ton kilometers (henceforth btkm) in 1951 to 82.36 btkm in 1971 (a 680 percent increase), and to 899.26 btkm in 2001 (a 1092 percent increase between 1971 and 2001) (Chaudhury (2005)). The rail freight volume also increased but at a much lower rate: from 127 btkm in 1971 to 312 btkm in 2001 (a 246 percent increase). Similar trends were observed for passenger traffic. This resulted in a dramatic reversal in the share of roads vs. railways: from 25 percent in 1951 to 75 percent in 2001 in favor of roads and highways.[14] This evidence motivates our measure of market access which is based on travel time through roads and highways. Note, however, that the estimates using colonial railroad as a source of identification may pick up some of the effects of the railroad to the extent colonial railroad length in the 1880s in a district is positively correlated with the railroad length in 1996 (or 2004). We underscore here that this poses no complications for our analysis because our goal is not to isolate the effects of roads and

[14]These estimates are from Chaudhury (2005). Alternative estimates reported by the Department of Roads and Highways of Government of India suggest a similar picture. In 1951, the share of roads and highways in freight traffic in India was 13.8 percent and in passenger 15.4 percent, which grew to 65 percent (freight) and 86.7 percent (passenger) in 2004-2005. While the national highways constitute about 2 percent of the total road network, it carries 40 percent of total road traffic (Annual Report, Department of Road Transport and Highways, GOI, 2006-2007).

highways from that of railroads, but to understand the effects of better market access on land inequality.

The Indian government invested heavily in transport infrastructure in the last few decades, with the expansion of Golden Quadrilateral network being one of the most ambitious projects. A number of interesting recent studies analyze the effects of this expansion and upgrading of the GQ network during the 2000s (see the discussion on related literature below). Note that our analysis does not attempt to estimate the effects of the recent expansion (6 lanes) and improvements in the GQ network. Because the effects of changes in trade costs due to the GQ expansion and upgrading on land inequality are likely to take many decades to materialize. Our analysis focuses on the fact that many parts of the GQ network have been in existence for a long time, and constituted the main transport arteries for long distance trade even before the Mughal period. The Grand Trunk Road which forms a large part of the Kolkata-Delhi arm of the GQ network is a prominent example, which goes back to Maurya era and underwent substantial improvements during the British rule, between 1833 and 1860 (Arnold (2000)). Our analysis thus deals with the long-term cumulative effects of better access to markets for the districts that are located closer to the different arms of the GQ network (for details on the identification scheme based on the GQ network, see section (4) below).

(3) Related Literature

Our analysis contributes to a large and growing literature on the effects of lower trade costs due to transport infrastructure investments. Donaldson (2015) and Berg et al. (2016) provide excellent surveys of this literature. Evidence suggests that a better access to markets reduces impediments to trade and spatial price dispersion (Donaldson (2018), Duranton (2015), Aggarwal (2013), Jones and Salazar (2021)), changes composition of trade, employment, and pattern of specialization (Michaels (2008), Duranton et al. (2014), Blankespoor et al. (2017)), increases household consumption in villages (Emran and Hou (2013)), causes deindustrialization (Faber (2014)), counters the effects of deindustrialization by increasing agricultural productivity and welfare (Blankespoor et al. (2022)), accelerates technology adoption, and structural change in and commercialization of agriculture and the rural economy (Damania et al. (2017), Fafchamps and Shilpi (2003), Emran and Shilpi (2012)), induces spatial decentralization of economic activities (Baum-Snow et al. (2017)).

India has been a prominent case study in the recent literature on the effects of trade costs on prices, productivity, allocational efficiency, and household welfare. There has been a surge of interest in understanding the effects of the recent expansion and upgrading of the GQ network. Ghani et al. (2016) find evidence that the GQ expansion and upgrading improved the organization and efficiency of the manufacturing activities through sorting, scaling, and reallocation, by both the incumbents and the new entrants. Datta (2012) find that the firms located closer to the GQ benefited in the form of more efficient inventory management. Das et al. (2019) provide evidence that the GQ spurred financial depth in the districts along the GQ, especially in the districts with a relatively more developed financial sector before the expansion of the GQ. Abeberese and Chen (2021) find both firm productivity rises and product scope falls as a result of the connection with the GQ highway. Estimates from a model of internal trade with variable markups calibrated to the Indian manufacturing sector suggest that real income increased by 2.7 percent as a result of lower trade costs from the expansion and upgrading of the GQ (Asturias et al. (2018)). However, to the best of our knowledge, there are no studies that analyze the effects of market integration owing to lower trade costs on land inequality in India or any other country.

(4) Empirical Issues and Identification Strategy

We calculate a gravity measure of market access using travel time (in hours) in 1996, and population in a destination district in 1991 as an indicator of market size. Following Allen and Atkin (2016), our main market access measure uses a trade elasticity of 1.5.[15] For details on the construction of the gravity measure of market access, please see section OA.1 in the online appendix. As noted earlier, the focus on roads and highways reflects the fact that roads and highways have become the main modes of transportation in India for both goods transport and traveling needs.

To understand the empirical issues, it is useful to consider the following triangular empirical

[15] The conclusions in this paper are robust to alternative choices of the trade elasticity parameter, and travel time in different years. Please see the online appendix.

model for estimating the impact of market integration on land Gini (LGini):

$$ln(LGini)_j = \delta_0 + \delta_1 ln(MA)_j + \Gamma X_{1j} + \zeta_j$$
$$ln(MA)_j = \alpha_0 + \Phi X_{1j} + \nu_j$$

where $ln(LGini)_j$ is log of land Gini in 2012 and $ln(MA)_j$ is the log of market access based on travel time in 1996 in district j, and X_{1j} is a vector of variables observed by the researcher that determines market access of a district and also affects land inequality.

The central empirical challenge in understanding the effects of market access on land inequality is that the placement of transport infrastructure is not random but determined by government policy objectives. The objectives may vary over time with political change, for example, when political parties have sharply different policy agendas, and may differ from the goals pronounced by the politicians publicly. It is not possible to identify and gather data on many of the variables that went into the actual route choice, and as a result, a vector of variables X_{2j} is omitted and subsumed in the error terms in the triangular model. This implies that $Corr\left(\zeta_j, \nu_j\right) \neq 0$. If the overriding objective for the government was to integrate the poor lagging regions to the growth centers, then the OLS estimate of the effects of transport infrastructure may be negative even though the true effect is positive when the poor regions have lower land inequality to begin with. Evidence on India, in fact, suggests that land inequality is lower in a poor district; a bivariate regression of $ln(LGini)$ on $ln(GDP)$ at the district level yields a coefficient of 0.01 with a t statistic of 7.7.

In contrast, when the roads are primarily targeted to areas with high economic potential to maximize economic growth and tax revenue, the OLS estimate of the better market access on an economic outcome is biased upward (towards a substantial positive effect) because these areas also have higher land inequality due to factors unrelated to market access. It may not be possible to pin down the net direction of bias arising from such endogeneous placement because, in general, both poverty targeting and tax revenue extraction have been important motives for governments and the objectives change over time. To address the biases in the OLS estimates, we exploit two sources of exogeneous variations in the market access of a district to develop an instrumental variables approach: (i) the location of colonial railroads in the 1880s, and (ii) the distance of a rural district from the Golden Quadrilateral highways. We discuss the credibility of these identifying sources in detail below.

(4.1) Colonial Railroads in the 1880s

The length of colonial rail track in the 1880s in a district is our first instrument for identifying the effects of market access of a district in 1996 on land inequality in 2012. We first discuss the plausibility of the exclusion restriction imposed. Then we explain why the colonial railroads are expected to be positively correlated with the market access in 2012 and report the relevant evidence.

As noted by Donaldson (2018), the locations of railroads built by the colonial government up to the 1880s were primarily dictated by defense considerations rather than economic objectives. The railroads were built and maintained by the military engineering core (National Transport Development Committee Report, vol 3, 2013, GOI). Following Donaldson (2018), we exclude the rail stations built after the 1880s as the Famine Commission report prompted the colonial government to target railroads to poor drought-prone districts more vulnerable to famine, thus making them potentially endogenous.

Some of the colonial railroads were financed by private railroad companies (Macpherson (1955)), and one might worry that they are likely to target the districts with higher economic potential to ensure adequate returns to the investors.[16] However, in colonial India, the privately financed railroads were guaranteed a 5 percent return by the government which ensured that the location choices were not driven by the imperative of ensuring a reasonable return to the investors. As a result, many of the private railroads were built in economically lagging districts. Hurd (1983, P. 743) writes: " ..., many, if not most, of the unprofitable lines depended for their very existence upon the guarantee. Those earning less than 5 per cent included some of the lines in the north-west and in the Ganges valley, most of those in the Deccan, and all of the lines in Sind and south India. Thus, ... , had the guarantee not existed, it is unlikely that private capital would have invested in railways for large areas of India. These areas would, then, have had no rail service at all."

Another potential threat to the exclusion restriction is the possibility that the colonial

[16] British investors invested 95 million pounds between 1845-1875 (Macpherson (1955)).

railroads might have led to agglomeration and persistent effects. If historical rail stations created centers of commerce, they might lead to agglomeration economies and persistent growth effects even after rail transport became less important or train stations were abandoned. A substantial body of recent research on the economic history of India allays this concern. The colonial railroads in India were unique in that they did not affect long-term growth and structural transformation in any significant way, unlike historical railroads in many other countries (see Bogart et al. (2015)). Thus, the long-term direct effects of colonial railroads on a district are likely to be negligible once we condition on the market access of a district in 1996. As a conservative strategy, we control for 1961 population density in a district to absorb any potential long-term impacts working through the agglomeration channel.[17]

A concern for the interpretation of the estimates based on the colonial railroads is that they might be picking up the effects of the colonial land revenue policies. In a widely cited paper, Banerjee and Iyer (2005) provide convincing evidence that the districts under the landlord system of land taxation had higher land inequality and lower irrigation investment. If the railroad length in a district is correlated with whether it was under the landlord system, then the IV estimates using the colonial railroads will partly reflect the effects of the colonial land revenue system. We will check this possibility by including an indicator for the colonial revenue system in a district using data from Banerjee and Iyer (2005).

The discussion above suggests that the exclusion restriction imposed on the colonial railroad is plausible. However, a reader might wonder whether we could be unaware of some other channels through which 1880s railroads may have very small direct effects on land inequality in 2012. Note that if this direct effect captures the role of the current railway network, we do not consider this as a violation of the identifying assumption. To the extent our instrument captures some of these other components of transport infrastructure (railway and water transport), it is part of the causal effect under focus. The sources of violation of the exclusion restriction have to be something different from these other transport infrastructures captured by the instrument. The important question here is whether allowing for such arbitrarily small direct impact of colonial railroad through non-market access channels have substantial im-

[17]Population density is the most commonly used indicator of agglomeration in economic geography.

pacts on the magnitude of the estimated causal effect of interest. To assess the sensitivity of the IV estimates with regards to such local (small) violation of the exclusion restriction, we take advantage of the Conley et al. (2012) bounds approach (see section 6.3 below).

The next question we address is that of relevance of the historical railroad as an instrument. Recall that our measure of market access is based on travel time through roads and highways. A natural question then arises: why should we expect colonial railroad to have a significant correlation with the gravity measure of market access based on roads and highways? If colonial railroad is only tangentially related to the market access through roads and highways in 1996 (or 1988, 2004) then the IV estimates will be biased and unstable, and can yield implausible magnitudes (Stock and Yogo (2005)).

To check whether the historical railroad locations have systematically higher market access, we plot the kernel density function of market access for two samples, with and without a colonial rail station. Figure 1 shows clearly that the presence of a colonial railroad in the 1880s shifts substantially the density function of market access in 1996 to the right. The districts with colonial railroad have a higher mean and lower variance.[18] The first stage F statistics later in the IV regressions confirm that the 1880 railroad length in a district has substantial power in explaining the variation in market access of districts in 1996.

There are a number of plausible reasons behind a significant positive correlation between colonial railroads and current market access via roads and highways in Figure 1. To the extent transport infrastructure placement is determined by topography, we would expect a positive correlation between the placement of rail line and roads. Two topographical features are especially important for the placement of railroads: slope and curvature. The optimal rail track location tries to minimize the slope (for steep slope, going up hill requires more powerful locomotive, and braking is difficult going down hill), and curves (a sharp curve reduces the maximum speed) (AREMA, 2003, Chapter 6). These two factors are also important for the choice of cost-minimizing road and highway routes.[19] Because of the topographical constraints,

[18] Districts with colonial railroad: mean=15.641 and variance=0.348. Districts without railroad: mean=14.911 and variance=0.372.

[19] The most widely used HDM highway planning model of the World Bank highlights the importance of slope (rise and fall) in choosing an "optimal' path for highways. See the discussion by Robinson and Thagensen (2004).

it is common to have rail tracks and highways placed close to each other. In fact, the old railroad bed may be the lowest-cost route for a new highway. As Duranton and Turner (2012) note: " (B)uilding both railroad tracks and automobile roads requires leveling and grading a roadbed. Hence, an old railroad track is likely to become a modern road...without the expense of leveling and grading."

(4.2) Golden Quadrilateral: An Inconsequential Place Design

The basic insight behind the inconsequential place design is that most of the interstate (national) highways are built to connect major metropolitan cities, and whether a village (or a small town) is located close to such highways is purely accidental and can be treated as quasi random (see Redding and Turner (2015) and Donaldson (2015) for excellent discussions).[20] In India, the Golden Quadrilateral highways (GQ) that connect 5 metropolitan cities: New Delhi, Kolkata, Chennai, Mumbai and Bangalore offer an excellent opportunity to develop an inconsequential place design (see Figure 2 for a map of the GQ network). For example, the fact that the distance from Patna (in the state of Bihar) to the Delhi-Kolkata arm of GQ is much lower than that from Darjiling (in the state of West Bengal) is not because GQ was targeted to Patna; the better exposure to markets for Patna is incidental (see Figure 2). We rely on such incidental variation in market access of different rural districts to identify the effects of market access on land inequality.

As noted earlier, many parts of the GQ network existed for centuries. A comparison of the roads and highways network in 1872 (Figure AF.1 in online appendix) with the network in 1992 in Figure 2 shows substantial overlaps. Most notably, the Grand Trunk Road (Kolkata to Delhi) goes back to ancient times and formed the main conduit for commerce and development ("the river of life" in the words of poet Rudyard Kipling) over centuries.[21] Our analysis attempts to capture the long term cumulative effects of market integration on land inequality across districts.

[20] For applications of inconsequential place design see, among others, Banerjee et al. (2012), Faber (2014), Datta (2012), and Ghani et al. (2016).

[21] The full length of this ancient transport corridor stretches as far north as Kabul, Afghanistan and as far south as Chittagong, Bangladesh. During the British rule, the Grand Trunk Road was developed into a two lane carriage and motor way (1833-1860). For discussions on the history of Grand Trunk Road in India, please see Singh (1995), and World Bank (2018).

For a credible empirical design, we need to address three issues in this set-up. First, as widely noted in the literature (see, for example, Faber (2014), Ghani et al. (2012)), we need to exclude the nodal districts (for example, Kolkata) as they were the targets of the GQ network. The nodal districts are cities, and thus are not in our sample given that the focus here is on rural land inequality. Second, the arms of GQ network show a substantial amount of zigzag, and one might worry (with some measure of justification) that the actual placement of the GQ arm reflects government targeting and political lobbying. Third, similar (perhaps stronger) concerns apply to the placement of feeder roads that connect a district center to the nearest point of the GQ arm. To strengthen the credibility of the identifying assumption, we need to purge such potentially endogeneous components of the GQ arm and the feeder roads. We implement an approach developed by Faber (2014) where in place of the actual road and highway network, a hypothetical road and highway network is used to purge out the endogeneous components.[22]

To construct the hypothetical highways and feeder roads, we use two approaches: (i) Euclidean distance, and (ii) the least cost path that exploits topographical features, especially slope and elevation. The Euclidean distance does not take into account the exogenous variation in the distance due to topographical constraints, for example, when the least cost path goes around a mountain rather than over it (or through it by tunnel). The deviations in least cost path from the linear (Euclidean) network arising from differences in elevation and slope provide a source of exogeneous variation in market access.[23] Our main results are based on the Euclidean distance and the corresponding estimates using the least cost path are reported in the online appendix. The advantage of the Euclidean distance is that it is independent of topographical features, and thus the exclusion restriction is not threatened by any direct impact of topography on land inequality. As an additional precaution, we include mean slope and elevation of a district as controls in all regressions in addition to other indicators of natural agricultural endowment such as rainfall (mean and SD) and an index of land productivity. However, these controls are more important for the exclusion restriction imposed

[22]See the discussion by Donaldson (2015) on the Faber (2014) approach.

[23]The insight that deviations of roads from a linear network caused by topography offers credible identifying variation has been used by many papers, for example, Emran and Hou (2013).

on the instrument based on the least cost path distance of a district to the nearest GQ arm.

(5) Data and Variables Definitions

Our analysis uses data constructed from several sources, which are presented below. Our unit of observation is the District defined from the 2001 census. Data on our outcome variables, land inequality, landlessness, modern technology adoption, come from the second round of the India Human Development Survey (IHDS) in 2012. The IHDS is a high quality household survey with a nationally representative coverage. IHDS II (2012) surveyed 42,152 households in 1,420 villages and 1,042 urban neighborhoods. IHDS was jointly organized by the University of Maryland and the National Council of Applied Economic Research in New Delhi (Desai et al., 2005; Desai and Vanneman, 2012). The summary statistics for the variables used in our analysis are reported in online appendix Table A.8.

Our main indicator of land inequality is the Gini coefficient for land ownership in a district. Formally, we calculate the Gini as follows:

$$= \frac{1}{2N^2\overline{y}} \sum_{i=1}^{N} \sum_{j \neq i}^{N-1} |y_i - y_j|$$

where: N is the number of households within the District; y_i is land owned by household i; $\overline{y}$ is the average land ownership within the District. As a second indicator of land inequality, we also consider the percentage of households in a district that are landless. Both measures of land inequality are calculated from the "area of land owned" variable of the IHDS survey.

To explore the role of technology adoption as a mechanism a' la Braverman and Stiglitz (1989), we use the share of households within a district that report using modern technology in agriculture. Specifically, whether they report using any one of the following equipment: tube well, electric pump, diesel pump, tractor/tiller, or a thresher. We also look at whether market integration helps deepen the formal credit market with an indicator of formal bank branch in a rural district. The survey has information on land sales at the household level, and we estimate the impacts of market integration on land sales.

We include several geographic control variables in our analysis. Climatological variables, including rainfall and temperature variation, are from BioClim and use 1961-1990 as reference (Hijmans et al., 2005). Elevation data are from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Data 30m (Farr et al., 2007). Mean slope estimates are from Verdin et al. (2007). Crop suitability is calculated as the maximum suitability among four high-input, rainfed crops (cotton, dry-land rice, maize, and wheat). These are calculated from Global Agro-Ecological Zone (GAEZ) data available from the Food and Agriculture Organization (Fischer et al., 2012). We also control for population density in 1961 using population data from the India District Database (Vanneman and Barnes, 2000). District area data are from Statoids.

For the GQ based identification scheme, we construct a hypothetical linear network connecting the main cities targeted by the GQ (see Figure 2). We then calculate the Euclidean distance between each district's centroid and this linear network. For robustness, we also consider the Euclidean distances to the least cost path GQ arms in the online appendix.

Our treatment variable, market access, is calculated as the weighted average of the populations of all other locations, with a weight that decreases with travel time. We calculate market access as follows: $MA_{it} = \Sigma_{i \neq j}(1/tt_{ijt}^{\theta})P_{jt}$, where MA_{it} = market access of District i at time t, tt_{ijt} is the travel time (in hours) between Districts i and j at time t, P_{jt}= population in destination District j at time t, and θ = trade elasticity.

Travel time between pairs of districts are from Allen and Atkin (2016). District populations for 1961 and 1991 are from Brinkhoff (2020). For trade elasticity, we follow Allen and Atkin (2016) and adopt 1.5 for our main results (and use alternative values for robustness). Our main results focus on market access calculated from travel time in 1996 and population in 1991.

(6) Evidence on the Effects of Market Integration on Land Inequality

We use two measures of land inequality for our empirical analysis: a land Gini at the district level and the proportion of landless households in a district. These measures are calculated for the year 2012. The regressions reported below include state fixed effects. State fixed effects are important for our empirical approach because there are important inter-state differences in land policy which can be traced back to the Zamindari system of revenue collection under British colonial rule (Banerjee and Iyer (2005)). The implementation of land reform in the

1960s and 1970s also varied substantially across different states (Ghatak and Roy (2007)). The state fixed effects also control for state-specific regulations of land market transactions. We build the evidence in a step by step fashion. The goal is to use a battery of alternative approaches and see if the evidence, taken together, leads to a robust conclusion.

(6.1) OLS and Oster Bias-Adjusted OLS Estimates

We begin with the OLS estimates, and check whether the OLS estimates remain robust once we correct for omitted variables bias using the approach developed by Altonji et al. (2005) and Oster (2019) where selection on observables is used as a guide to selection on unobservables. In particular, we use the bias adjusted OLS estimator (henceforth BA-OLS) proposed by Oster (2019). The advantage of the BA-OLS estimator is that the conclusions do not rely on any exclusion restrictions, but, unlike the IV estimates, this approach cannot correct for attenuation bias due to measurement error in the market access measure.

The estimates from OLS and Oster (2019) BA-OLS estimators are reported in Table 1. The OLS estimate of the coefficient of the indicator of market access ($ln(MA)_j$) is 0.108 without any controls, and it is significant at the 1 percent level. To check whether the positive impact of market integration found in the baseline OLS estimate is driven by unobserved heterogeneity, we include state fixed effects and a set of agro-climatic controls that can affect the productivity of land: an index of crop suitability (from FAO), the mean elevation and slope of a district, the long-term average and standard deviation of rainfall, the average of and seasonality in temperature.[24]

The estimates in column 2 of Table 1 are striking; the point estimate remains virtually unchanged (0.108 (column 1) to 0.111 (column 2)), even though the R^2 more than doubles from 0.166 (column 1) to 0.380 (column 2) once we add the control variables. As discussed by Oster (2019), the sensitivity of an OLS estimate to the inclusion of control variables is informative about the importance and direction of omitted variables bias only when the control variables increase the R^2 substantially. The BA-OLS estimate in column 3 of Table 1 corrects for selection on unobservables in addition to the observed control variables added in column 2,

[24] Table 1 does not report the controls in the regressions. Please see online appendix Table A.1 for the full Table.

and the point estimate again increases slightly: from 0.110 (column 2) to 0.115 (column 3). This pattern of estimates contradicts the idea that the unobserved heterogeneity biases the estimated impact of market integration upward. This suggests that the OLS estimate with controls is likely to be biased downwards when we take into account measurement error in the measure of market access $ln(MA)_j$.

The OLS and BA-OLS estimates of the effects of market integration on the proportion of landless households also suggest a positive impact of market integration which is significant at the 1 percent level across the board (see columns 4-6 in table 1). The estimated coefficients vary only marginally: from 0.360 (OLS without controls) to 0.357 (OLS with controls) to 0.352 (BA-OLS). Again this lack of sensitivity is observed despite the fact that the set of controls have substantial explanatory power: the R^2 increases from 0.167 to 0.428 once we add the controls. This strengthens the idea that the OLS estimates of the effects of market integration on land inequality are biased downward.

(6.2) IV Estimates of the Impact of Market Integration on Land Inequality

The IV estimates are reported in Tables 2A (land Gini) and 2B (landless).[25] The regression specification includes the set of controls in column 2 of Table 1 discussed above in addition to the state fixed effects. We report 3 different estimates for each measure of land inequality. The first two columns use 2SLS and rely on the 1880 railroad length (column 1) and Euclidean distance to the hypothetical linear GQ network (column 2) as identifying instruments. The third column reports estimates from Lasso-IV based on an extended set of instruments consisting of colonial railroad, Euclidean distance to the nearest GQ arm, and the interactions of these two instruments with all the exogeneous controls in the model such as slope, elevation, and rainfall. The Lasso-IV picks a parsimonious subset of efficient instruments from the extended set of instruments.

The evidence suggests that the 1880 railroad is a particularly strong source of exogeneous variation in our market access variable with first stage F statistics of 18.94 (land Gini regression) and 18.49 (landless regression). Euclidean distance to linear GQ network has good

[25] Again, for the sake of brevity, we do not report the control variables in Table 2A and 2B. The full Tables including the controls are reported in the online appendix. Please see Tables A.2A and A.2B.

power in the landless regression with an F statistic of 11.14, but it has moderate power in the land Gini regression (F=6.25). Recent advances in the literature on the properties of just-identified IV estimate when the IV lacks power is helpful in this context. In an interesting recent analysis, Angrist and Kolesár (2023) provide evidence that the weak instrument bias is low in most microeconometric applications for the just-identified model.[26] We thus expect the estimates from the GQ IV used in a just-identified model to be credible.

Consistent with a priori expectations, a district with longer railroad network in the 1880s has better market access in 1996, and a district further away from the linear GQ network has a lower market access, and both instruments are significant at the 1 percent level irrespective of the indicator of land inequality. Since Lasso picks multiple instruments, we can implement the estimation in two different ways. The first and straightforward approach is to use the set of instruments in a 2SLS regression. However, given the evidence that weak instrument bias in a just-identified model is small, we can exploit this by adding a zero stage that predicts market access using the set of instruments along with all other exogeneous variables and then use the predicted MA as a single instrument. This procedure converts a over-identified model into a just identified model.[27] As noted by Kolesár et al. (2015), this approach relies on a weaker identifying assumption in that we do not impose exclusion restrictions separately on each instrument.[28] However, the point estimates from these alternative approaches are very close in all our estimation. Column (4) in Tables 2A and 2B reports the estimate from Lasso-IV using the multiple instruments directly. The last column (5) contains estimate from a specification where we combine the three interaction based instruments picked by lasso in column (4) with the railroad and GQ instruments and use the predicted market access from a zero stage to convert it to a just identified model. The point estimates vary somewhat between

[26] Angrist and Kolesár (2023) write "... in microeconometric applications, just-ID IV estimators can typically be treated as all but unbiased and that the usual inference strategies are likely to be adequate".

[27] For application of this procedure, see Rajan and Subramanian (2008), and Emran et al. (2020), among others.

[28] Note that we do not report Hansen's J test as a test for validity of the exclusion restrictions when using both historical railroad and GQ instruments. Since there is no reason to expect that the compliers for the two instruments overlap substantially, the IV estimates from just identified models using the alternative instruments one at a time should be different. In fact, the estimates in table 2 show clearly the heterogeneity (compare the results in column 4 and 5 for landless in Table 2) and a Hansen's J test would reject the exclusion restriction incorrectly because of heterogeneity in the effects of market integration.

columns (4) and (5) for land gini (Table 2A), but are very close for landless (Table 2B).

The IV estimates strengthen the conclusion that market integration increases land inequality, the impact on both land Gini and proportion of landless households in a district is positive and significant at the 1 percent level. The numerical magnitudes of the estimates are larger compared to the corresponding Oster (2019) BA-OLS estimate. This probably reflects a combination of correction of attenuation bias and dealing with the omitted variables bias in a more adequate way. To have sense of the magnitudes of the impacts, we focus on the Lasso-IV estimates. The Lasso-IV estimates imply that land Gini in a district with 10 percent higher market access is 2.6 percent higher. For the impact on the landless, the corresponding estimate is a 6.78 percent higher landlessness. These are clearly substantial impacts.

(6.3) Relaxing the Exclusion Restriction: Evidence from Conley et al. (2012) Bounds

The IV estimates in Table 2A and 2B require that the exclusion restriction imposed on an instrument holds *exactly* (Conley et al. (2012)), i.e., the IV has a precisely zero direct impact on the outcomes of interest: land Gini and the proportion of landless in a district. A reader might worry that this "exact exclusion restriction" may be violated locally where an instrument exerts a small direct impact (positive or negative) through some unspecified channels. It is thus a reasonable question to ask: is the main conclusion that market integration increases land inequality robust to allowing for such small direct impact of the instruments?

To address this, we implement the approach developed by Conley et al. (2012). The relaxation of the exact exclusion restriction implies that we no longer have point identification, but can estimate bounds on the causal effect of interest. To understand the basic intuition behind the approach, consider the following extension of the triangular empirical model set out earlier in section 3 above (with colonial railroad (denoted as R_j) as the identifying instrument):

$$ln(LGini)_j = \delta_0 + \delta_1 ln(MA)_j + \Gamma X_{1j} + \theta R_j + \zeta_j$$
$$ln(MA)_j = \alpha_0 + \Phi X_{1j} + \beta R_j + \nu_j$$

The IV (2SLS) estimates in Tables 2A rely on the following identifying assumptions: $\theta = 0$ (exact exclusion restriction) and $\beta \neq 0$ (instrument relevance). Conley et al. (2012) develop methods to estimate bounds on the parameter δ_1 under the assumption that θ belongs to a narrow interval around zero, i.e., $\theta \in [-\epsilon, +\epsilon]$ for arbitrarily small values of $\epsilon > 0$. In particular, we implement the "UCI" (union of confidence intervals) method proposed by Conley et al. (2012). This approach is the most conservative as it only specifies the support of the distribution for the parameter θ.

The results from the Conley et al. (2012) bounds approach are reported in Table 3. We report bounds on the estimated δ_1 for three values of $\epsilon = 0.0001, 0.001, 0.01$. The results show that the estimated bound for the parameter remains positive even when we assume a relatively large interval for θ with $\epsilon = 0.01$, except for the cases when Euclidean distance to GQ is used as the sole instrument. The wide bounds for the GQ instrument reflect its lack of strength in the first stage regression discussed earlier.[29] The estimates from the IVs picked by Lasso together provide us the most credible evidence and the impact of market integration on both measures of land inequality (land Gini and proportion of landless) remains numerically substantial.

(6.4) Robustness Checks

The main results on land inequality in 2012 presented in Tables 1-3 are based on a measure of market access calculated using travel time in 1996 (based on Allen and Atkin (2016)). We check whether the conclusions change when we use travel time from other years. We use 2004 travel time to calculate market access, and the results from alternative estimators are reported in Table A.3 in the online appendix. The estimates are broadly consistent with the conclusion that market integration increases land Gini and the proportion of landless in a district.

We also provide evidence on potential sensitivity of the main conclusions to different assumptions regarding the trade elasticity parameter in calculating market access. As discussed before, our main results are based on a trade elasticity of 1.5. In online appendix Table A.4, we report estimates for an alternative value based on Donaldson (2018): 3.8. The results are again consistent with the main conclusions based on Tables 1, 2A, and 2B in the paper.

[29]We, however, emphasize that, consistent with Angrist and Kolesár (2023), the estimated effects from the just identified model with the GQ instrument are close to the other estimates without any weak instrument issue (please see Tables 2A and 2B). This suggests that the estimates using solely the GQ instrument for identification provides credible estimates.

Next, we check if the positive impacts of market integration on land inequality in Tables 1, 2A, and 2B partly capture the effects of (i) differences in colonial land revenue system, (ii) demographic pressure, and (iii) differences in inheritance rules (laws and customs) between Hindu and Muslim populations. As noted earlier, Banerjee and Iyer (2005) provide evidence that colonial land tax policies had long-term effects on land inequality in India. To check if our IV estimates (especially using the colonial railroad as a source of identifying variation) partly captures the persistent effects of land revenue system, we include two dummies indicating the type of land tax system was in place in a district during the British colonial period. The estimates in column (1) of Table A.5 in the online appendix suggests that our estimated effects of market access remain virtually unchanged.

In an analysis of land inequality in West Bengal, Bardhan et al. (2014) show that land inequality is influenced by the demographic changes through population growth, and land inheritance law and customs. Note that we include 1961 population density in all the regressions as a control for possible agglomeration effects of historical infrastructure. This also takes care of differences in population growth and demographic pressure on land up till 1961. To understand the role of population growth after 1961, we include 2011 population density as an additional control. We also include the proportion of Muslim in the regressions and find that the impact of market integration on land inequality is barely affected by the inclusion of these two control variables (see online appendix Table A.5).

(7) Mechanisms

The theoretical analysis of Braverman and Stiglitz (1989) emphasize the role of technology adoption as a mechanism for increasing land inequality in response to market integration caused by lower trade costs. Since agricultural technology that can give rise to increasing returns is especially important, we estimate the impact of market integration on a measure of technology adoption based on the ownership of the following farming equipments: tube well, electric pump, diesel pump, tractor/tiller, or a thresher. We also check whether market integration has had a significant effect on land sales in a district in 2012. Table 4 reports the estimates for technology adoption (panel A) and land sales (panel B). We also report estimates of the effects on financial deepening measured by the presence of a formal bank branch (see panel C). For each outcome of interest, we present 8 estimates, including OLS, BA-OLS, and different IV estimates.

The evidence suggests that market integration increases adoption of farming technology that are subject to increasing returns, and the effect is statistically significant at the 1 percent level in the specification using the Lasso-IVs (see column (7) in Table 4). The magnitude of the effect is also not small: a 10 percent increase in the market access index increases the adoption of technology by 3.5 percent. The evidence on land sales is statistically not precise, but the estimates overall suggest a positive impact of market integration. Even though the OLS estimate of the effects on land sales is positive, numerically substantial, and significant at the 5 percent level after controlling for the agroclimatic heterogeneity and 1961 population density, the BA-OLS and IV estimates have large standard errors. Interestingly, the point estimates from BA-OLS and Lasso-IV are larger in magnitude when compared to the OLS estimates in columns 1 and 2 (see panel B of Table 4).

When market integration deepens the formal credit market through expansion of bank branches, the advantages enjoyed by the large land owners relative to the functionally landless and small landholders would be reinforced. Das et al. (2019) provide an extensive analysis providing convincing evidence that better market access does in fact help develop the formal financial sector in India. We also add some suggestive evidence. Panel C in Table 4 report estimated impacts of market integration on the access to formal banks: the dependent variable of interest being a dummy for the existence of a formal bank branch in a village. The estimates suggest that a better market access leads to a higher probability of having a formal bank branch in a village. The evidence taken together thus suggests that the deepening of the formal financial sector is an important mechanism for understanding the effects of market integration on land inequality.

(8) Conclusions

The world witnessed dramatic improvements in transport infrastructure in the last few decades, which substantially reduced trade costs and led to spatial market integration. We provide evidence on the effects of the market integration on land inequality in the rural areas. Our empirical analysis uses data on land ownership from a high quality household survey in India, and exploits two sources of exogeneous variation: a historical infrastructure design based on colonial railroads in the 1880s and an inconsequential place design based on the Golden Quadrilateral network of highways. We also report estimates from the Oster (2019) approach that does not impose any exclusion restrictions, and following Altonji et al. (2005), relies on selection on observables as a guide to unobservables for tackling the omitted variables biases.

The evidence suggests that market integration increases land inequality in a district: a 10 percent higher market access leading to a 2.6 percent increase in land inequality (land Gini), and a 6.8 percent increase in the incidence of landless. These conclusions are robust across alternative econometric approaches, and different measures of market access. The conclusion that market integration increases land inequality holds even when we relax the exclusion restrictions imposed in the IV estimation by using the "plausibly exogenous" approach developed by Conley et al. (2012).

We explore the mechanisms behind the observed impact of market integration on land inequality. We find evidence that market integration increases the adoption of increasing returns technology in agriculture, a mechanism emphasized by the theoretical model of Braverman and Stiglitz (1989). The evidence on the effects of market access on land sales shows a positive impact, but the estimates are imprecise. Evidence also suggests that market integration leads to a deepening of the formal banks, which reinforces the advantages of the large landholders in the land market transactions.

References

Abeberese, A. B. and Chen, M. (2021). Intranational trade costs, product scope and productivity: Evidence from india's golden quadrilateral project. *Journal of Development Economics*, page 102791.

Aggarwal, S. (2013). Do rural roads create pathways out of poverty. *Evidence from India: UC Santa Cruz.*

Allen, T. and Atkin, D. (2016). Volatility and the Gains from Trade. NBER Working Papers

22276, National Bureau of Economic Research, Inc.

Altonji, J., Elder, T., and Taber, C. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1):151–184.

Angrist, J. and Kolesár, M. (2023). One instrument to rule them all: The bias and coverage of just-id iv. *Journal of Econometrics*.

Arnold, D. (2000). The New Cambridge History of India. Cambridge University Press.

Asturias, J., GarcÃa-Santana, M., and Ramos, R. (2018). Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India. *Journal of the European Economic Association*, 17(6):1881–1940.

Atkin, D. and Donaldson, D. (2015). Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs. NBER Working Papers 21439, National Bureau of Economic Research, Inc.

Banerjee, A., Duflo, E., and Qian, N. (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. In *NBER Working Paper*. National Bureau of Economic Research.

Banerjee, A. and Iyer, L. (2005). History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review*, 95(4):1190–1213.

Banerjee, A. V. and Newman, A. F. (1993). Occupational Choice and the Process of Development. *Journal of Political Economy*, 101(2):274–298.

Bardhan, P., Luca, M., Mookherjee, D., and Pino, F. (2014). Evolution of land distribution in West Bengal 19672004: Role of land reform and demographic changes. *Journal of Development Economics*, 110:171–190.

Bauluz, L., Govind, Y., and Novokmet, F. (2020). Global Land Inequality. World Inequality Lab Working Papers halshs-03022360, HAL.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics*, 99(3):435–448.

Berg, C. N., Deichmann, U., Liu, Y., and Selod, H. (2016). Transport Policies and Development. *The Journal of Development Studies*, 53(4):465–480.

Besley, T. and Burgess, R. (2000). Land reform, poverty reduction, and growth: Evidence from india. *The Quarterly Journal of Economics*, 115(2):389–430.

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

Bharti, N. (2018). Wealth Inequality, Class and Caste in India, 1961-2012. Working paper no. 14, WorldInequality Lab.

Binswanger, H. P., Deininger, K., and Feder, G. (1995). Power, distortions, revolt and reform in agricultural land relations. In Chenery, H. and Srinivasan, T., editors, *Handbook of Development Economics*, volume 3 of *Handbook of Development Economics*, chapter 42, pages 2659–2772. Elsevier.

Blankespoor, B., Bougna, T., Garduno Rivera, R., and Selod, H. (2017). Roads and the geography of economic activities in mexico. *World Bank Policy Research Working Paper*, (8226).

Blankespoor, B., Emran, M. S., Shilpi, F., and Xu, L. (2022). Bridge to Bigpush or Backwash? Market Integration, Reallocation, and Productivity Effects of Jamuna Bridge in Bangladesh. *Journal of Economic Geography.*

Bogart, D., Chaudhary, L., and Herranz-Loncan, A. (2015). The Growth Contribution of Colonial Indian Railways in Comparative Perspective. CEH Discussion Papers 033, Centre for Economic History, Research School of Economics, Australian National University.

Braverman, A. and Stiglitz, J. (1989). Credit rationing, tenancy, productivity, and the dynamics of inequality. In Bardhan, P., editor, *The Ecnomic Theory of Agrarian Institutions.* Oxofrd University Press.

Brinkhoff, T. (2020). *City Population http://www.citypopulation.de.*

Chaudhury, P. D. (2005). Modal split between rail and road modes of transport in india. *Vikalpa*, 30(1):17–34.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly Exogenous. *The Review of Economics and Statistics*, 94(1):260–272.

Damania, R., Berg, C., Russ, J., Barra, A. F., Nash, J., and Ali, R. (2017). Agricultural Technology Choice and Transport. *American Journal of Agricultural Economics*, 99(1):265–284.

Das, A., Ghani, E., Grover, A., Kerr, W., and Nanda, R. (2019). Infrastructure and Finance: Evidence from India's GQ Highway Network. Harvard Business School Working Papers 19-121, Harvard Business School.

Datta, S. (2012). The impact of improved highways on indian firms. *Journal of Development Economics*, 99(1):46–57.

Desai, S., Dubey, A., Joshi, B., Sen, M., Shariff, A., Vanneman, R., and Codebook, H. (2005). India human development survey (ihds). *NCAER, India and Inter-university Consortium for Political and Social Research, MI, USA*.

Desai, S. and Vanneman, R. (2012). India human development survey (ihds)-ii. icpsr36151. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research*.

Donaldson, D. (2015). The Gains from Market Integration. *Annual Review of Economics*, 7(1):619–647.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Duranton, G. (2015). Roads and trade in Colombia. *Economics of Transportation*, 4(1):16–36.

Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and Trade: Evidence from the US. *The Review of Economic Studies*, 81(2):681–724.

Duranton, G. and Turner, M. A. (2012). Urban Growth and Transportation. *Review of Economic Studies*, 79(4):1407–1440.

Emran, M. S. and Hou, Z. (2013). Access to Markets and Rural Poverty: Evidence from Household Consumption in China. *Review of Economics and Statistics*, 95(2):682–697.

Emran, M. S., Islam, A., and Shilpi, F. (2020). Distributional Effects of Corruption When Enforcement is Biased: Theory and Evidence from Bribery in Schools in Bangladesh. *Economica*, 87(348):985–1015.

Emran, M. S. and Shilpi, F. (2012). The extent of the market and stages of agricultural specialization. *Canadian Journal of Economics*, 45(3):1125–1153.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from Chinas National Trunk Highway System. *The Review of Economic Studies*, 108((4-5)):899–934.

Fafchamps, M. and Shilpi, F. (2003). The spatial division of labour in Nepal. *Journal of Development Studies*, 39(6):23–66.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al. (2007). The shuttle radar topography mission. *Reviews of geophysics*, 45(2).

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Tóth, G., Van Velthuizen, H., and Wiberg, D. (2012). Global agro-ecological zones (gaez v3. 0): Model documentation. international institute for applied systems analysis (iiasa), laxenburg.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. *The Economic Journal*, 126(591):317–357.

Ghatak, M. and Roy, S. (2007). Land reform and agricultural productivity in India: a review of the evidence. *Oxford Review of Economic Policy*, 23(2):251–269.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(15):1965–1978.

Himanshu (2019). Inequality in india. Technical report.

Jones, S. and Salazar, C. (2021). Infrastructure Improvements and Maize Market Integration: Bridging the Zambezi in Mozambique. *American Journal of Agricultural Economics*, 103(2):620–642.

Kanbur, R. and Venables, A. J., editors (2005). *Spatial Inequality and Development*. Oxford University Press.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., and Imbens, G. (2015). Identification and Inference With Many Invalid Instruments. *Journal of Business & Economic Statistics*, 33(4):474–484.

Mariscal, E. and Sokoloff, K. L., editors (2000). *Schooling, Suffurage and the Persistance of Inequality in the Americas, 1800-1945*. Palo Alto: Hoover Institution Press.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, 90(4):683–701.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Rajan, R. and Subramanian, A. (2008). Aid and Growth: What Does the Cross-Country Evidence Really Show? *The Review of Economics and Statistics*, 90(4):643–665.

Redding, S. and Turner, M. (2015). Transportation Costs and the Spatial Organization of Economic Activity. In *Handbook of Regional and Urban Economics*, pages 1339–1398. Elsevier.

Robinson, R. and Thagensen, B., editors (2004). *Road Engineering for Development*. Spon Press: Taylor and Francis Group.

Singh, R., editor (1995). *Grand Trunk Road: A Passage Through Time*. Aperture Books.

Sokoloff, K. L. and Engerman, S. L. (2000). Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives*, 14(3):217–232.

Tierney, M. J., Nielson, D. L., Hawkins, D. G., Roberts, J. T., Findley, M. G., Powers, R. M., Parks, B., Wilson, S. E., and Hicks, R. L. (2011). More Dollars than Sense: Refining Our Knowledge of Development Finance Using AidData. *World Development*, 39(11):1891–1906.

van Maarseveen, R. (2020). The urban-rural education gap: do cities indeed make us smarter? *Journal of Economic Geography*, 21(5):683–714.

Vanneman, R. and Barnes, D. (2000). Indian district data, 1961-1991: machinereadable data file and codebook. *Center on Population, Gender, and Social Inequality. University of Maryland*.

Verdin, K. L., Godt, J. W., Funk, C., Pedreros, D., Worstell, B., and Verdin, J. (2007). Development of a global slope dataset for estimation of landslide occurrence resulting from earthquakes. Technical report.

World Bank (2005). *World Development Report 2006*. Oxford University Press.

World Bank (2018). The Web of Transportation Corridors in South Asia. Technical report, World Bank Group, Washington DC.

Table 1. Effects of Market Access on Land Inequality:
OLS and Bias-Adjusted OLS Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.	**ln(Landgini 2012)**			**ln(Landless 2012)**		
Estimator	OLS	OLS	Bias-Adjusted	OLS	OLS	Bias-Adjusted
ln(MA1996)	**0.108***** **(0.02)**	**0.111***** **(0.02)**	**0.115***** **(0.04)**	**0.360***** **(0.07)**	**0.357***** **(0.07)**	**0.352***** **(0.11)**
Control variables	No	Yes	Yes	No	Yes	Yes
State Dummies	No	Yes	Yes	No	Yes	Yes
Observations	200	200	200	212	212	212
R-squared	0.166	0.380		0.167	0.428	

Notes: This table shows the estimated effects of market access in 1996 on land inequality in 2012. Market Access is calculated using population in 1991 and travel time in 1996. Columns (1), (2) and (3) measure land inequality using a Gini index in 2012. Columns (4), (5), and (6) measure land inequality as the share of landless households in a district. The unit of analysis is the district level according to the 2001 Census boundaries. We control for geographic variables (slope, elevation, crop suitability, rain, temperature, rain coefficient of variation, temperature seasonality), population density in 1961, and state fixed effects. Columns (1), (2), (4) and (5) are estimated by OLS. Columns (3) and (6) are estimated by Oster's bias-adjusted OLS (Oster (2019)). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The full Table with the controls is reported in the online appendix.

Table 2A. Effects of Market Access on Land Gini (IV Estimates)

	(1)	(2)	(3)	(4)	(5)
Dep. var.			**ln(Landgini 2012)**		
Estimator	2SLS	2SLS	2SLS	Lasso-IV	2SLS
ln(MA1996)	**0.244*** **	**0.232** **	**0.241*** **	**0.255*** **	**0.197*** **
	(0.07)	**(0.09)**	**(0.06)**	**(0.06)**	**(0.05)**
Control variables	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes
First Stage					
ln(km of railroad)	0.117***		0.114***		0.192***
	(0.03)		(0.03)		(0.07)
ln(dist. to GQ)		-0.128***	-0.120**		-0.015
		(0.05)	(0.05)		(0.08)
ln(dist. to GQ) x				-0.131***	-0.112*
ln(slope)				(0.04)	(0.06)
ln(dist. to GQ) x				0.152	0.106
Elevation				(0.18)	(0.19)
ln(km of railroad) x				0.002***	-0.001
Crop suitability				(0.00)	(0.00)
Angrist-Pischke F	**18.94**	**6.25**	**21.93**	**9.19**	**34.23**
	0.0000	0.0134	0.0000	0.0000	0.0000
Observations	200	200	200	200	200
R-squared	0.239	0.263	0.247	0.216	0.321

Notes: This table reports the estimated effect of market access on the land Gini index in 2012. Market Access is calculated using population in 1991 and travel time in 1996. We control for geographic variables (slope, elevation, crop suitability, rain, temperature, rain coefficient of variation, temperature seasonality), population density in 1961, and state fixed effects. Columns (1) and (2) are estimated by two-stage least squares (2SLS) using railroad length in the 1880s in a district, and distance of a district to the nearest arm of Golden Quadrilateral as instrumental variables, respectively. Column (3) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero stage with both railroad length and distance to GQ. The Lasso in column (4) is estimated using a parsimonious set of instruments chosen by Lasso from a broad set that include railroad length, distance to GQ, and each of their interactions with the exogenous control variables as instrumental variables. Lasso selects three instruments which are reported in the first stage in column (4). Column (5) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero-stage including railroad length, distance to GQ, and all the instruments chosen by Lasso. In columns (3) and (5), under the first stage, we report the zero stage estimated coefficients of the instrumental variables. The first stage reports the estimated coefficients of the instrumental variables from a regression of market access on the controls and instruments. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The full Table with the controls is in the online appendix.

Table 2B. Effects of Market Access on Landlessness (IV Estimates)

	(1)	(2)	(3)	(4)	(5)
Dep. var.			ln(Landless 2012)		
Estimator	2SLS	2SLS	2SLS	Lasso-IV	2SLS
ln(MA1996)	**0.974***	**0.403***	**0.758***	**0.678***	**0.662***
	(0.23)	**(0.21)**	**(0.16)**	**(0.14)**	**(0.14)**
Control variables	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes
First Stage					
ln(km of railroad)	0.114***		0.114***		0.186***
	(0.03)		(0.03)		(0.07)
		-			
ln(dist. to GQ)		0.148***	-0.120**		-0.030
		(0.04)	(0.05)		(0.08)
ln(dist. to GQ) x				-0.131***	-0.106*
ln(slope)				(0.04)	(0.06)
ln(dist. to GQ) x				0.119	0.081
Elevation				(0.17)	(0.18)
ln(km of railroad) x				0.002***	-0.001
Crop suitability				(0.00)	(0.00)
Angrist-Pischke F	**18.49**	**11.14**	**26.96**	**12.36**	**41.70**
	(0.0000)	(0.0010)	(0.0000)	(0.0000)	(0.0000)
Observations	212	212	212	212	212
R-squared	0.154	0.427	0.312	0.3535	0.361

Notes: This table reports the estimated effect of market access on the share of landless households in the district, in 2012. Market Access is calculated using population in 1991 and travel time in 1996. We control for geographic variables (slope, elevation, crop suitability, rain, temperature, rain coefficient of variation, temperature seasonality), population density in 1961, and state fixed effects. Columns (1) and (2) are estimated by two-stage least squares (2SLS) using railroad length and distance to the Golden Quadrilateral as instrumental variables, respectively. Column (3) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero stage with including both railroad length and distance to GQ. The Lasso in column (4) is estimated using a parsimonious set of instruments chosen by Lasso from a broad set that includes railroad length, distance to GQ, and each of their interactions with the exogenous control variables as instrumental variables. Lasso selects three instruments which are reported in the first stage in column (4). Column (5) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero stage with railroad length, distance to GQ, and the instruments chosen by Lasso. The predicted Market Access are fitted values from a zero-stage including railroad length, distance to GQ, and all the instruments chosen by Lasso. In columns (3) and (5), under the first stage, we report the zero stage estimated coefficients of the instrumental variables. The first stage reports the estimated coefficients of the instrumental variables from a regression of market access on the controls and instruments. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The full Table with controls is in the online appendix.

Table 3. Relaxing the Exclusion Restrictions: Conley et al. Bound Estimates

		Lower Bound	Upper Bound
ln(Landgini 2012)			
ln(km of railroad)	θ ϵ [-0.01*β, 0.01*β]	0.097	0.399
	θ ϵ [-0.05*β, 0.05*β]	0.023	0.502
	θ ϵ [-0.1*β, 0.1*β]	-0.082	0.636
ln(dist. to GQ)	θ ϵ [-0.01*β, 0.01*β]	0.028	0.445
	θ ϵ [-0.05*β, 0.05*β]	-0.033	0.541
Predicted Market Access 1	θ ϵ [-0.01*β, 0.01*β]	0.129	0.353
	θ ϵ [-0.05*β, 0.05*β]	0.121	0.364
	θ ϵ [-0.1*β, 0.1*β]	0.110	0.378
Predicted Market Access 2	θ ϵ [-0.01*β, 0.01*β]	0.140	0.371
	θ ϵ [-0.05*β, 0.05*β]	0.131	0.383
	θ ϵ [-0.1*β, 0.1*β]	0.120	0.398
ln(Landless 2012)			
ln(km of railroad)	θ ϵ [-0.01*β, 0.01*β]	0.441	1.549
	θ ϵ [-0.05*β, 0.05*β]	0.157	1.991
	θ ϵ [-0.1*β, 0.1*β]	-0.263	2.563
ln(dist. to GQ)	θ ϵ [-0.01*β, 0.01*β]	-0.068	0.876
Predicted Market Access 1	θ ϵ [-0.01*β, 0.01*β]	0.427	1.091
	θ ϵ [-0.05*β, 0.05*β]	0.402	1.124
	θ ϵ [-0.1*β, 0.1*β]	0.370	1.165
Predicted Market Access 2	θ ϵ [-0.0001, 0.0001]	0.375	1.022
	θ ϵ [-0.05*β, 0.05*β]	0.351	1.053
	θ ϵ [-0.1*β, 0.1*β]	0.320	1.092

Notes: This table reports the upper and lower bound estimates of the coefficient on market access under the assumption that the instrumental variable is "plausibly exogenous" (Conley et al. 2012). (1) θ is the direct effect of an instrument on the outcome variable. The lower and upper bounds are the estimated effects of market access on the relevant measure of land inequality given that θ belongs to a specified interval. (2) Predicted Market Access refers to estimated fitted values from a zero-stage regression. This allows us to estimate a just identified IV model using predicted market access as a single instrument. (3) Predicted Market Access 1 is from a zero stage with both railroad length and distance to GQ. Predicted Market Access 2 is from a zero stage with the three instruments chosen by the Lasso: distance to GQ interacted with slope, distance to GQ interacted with elevation, and railroad length interacted with crop suitability.

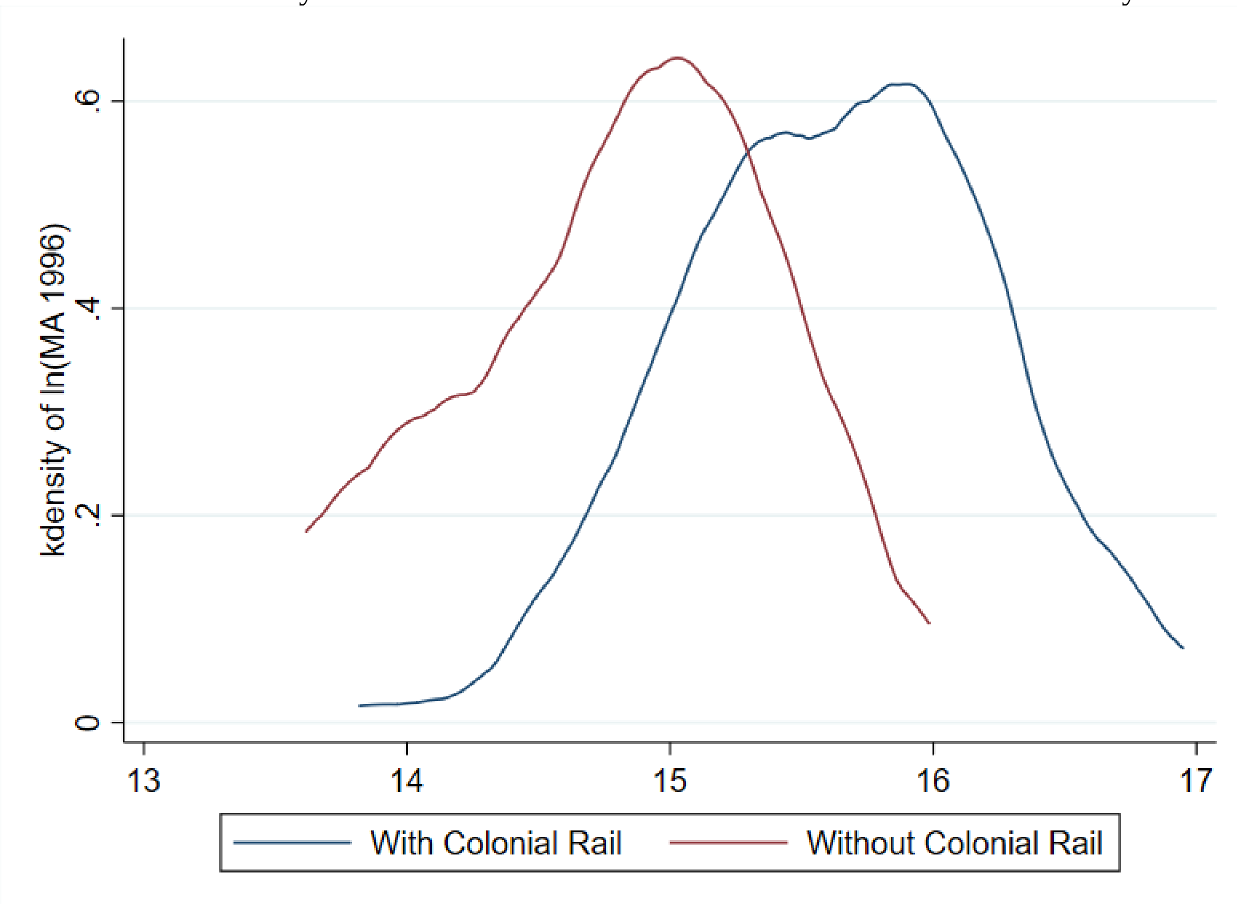
Table 4. Understanding the Mechanisms

Estimator	(1) OLS	(2) OLS	(3) Bias-Adj.	(4) 2SLS	(5) 2SLS	(6) 2SLS	(7) Lasso-IV	(8) 2SLS
Panel A				**ln(Tech. Use 2012)**				
ln(MA1996)	0.273***	0.235***	0.188	0.423***	0.289**	0.366***	0.350***	0.364***
	(0.06)	(0.04)	(0.70)	(0.14)	(0.12)	(0.09)	(0.09)	(0.08)
First Stage F				18.49	11.14	26.95	12.36	41.70
				0.0000	0.0010	0.000	0.0000	0.000
R-squared	0.186	0.538		0.488	0.533	0.5136	0.519	0.514
Panel B				**ln(Land Sale, 2012)**				
ln(MA1996)	1.172*	2.330**	3.093***	-1.537	1.902	-0.072	0.433	1.127
	(0.66)	(0.90)	(0.96)	(2.65)	(3.13)	(2.04)	(2.11)	(1.90)
First Stage F				18.49	11.14	26.95	12.36	41.70
				0.0000	0.0010	0.000	0.0000	0.000
R-squared	0.014	0.163		0.080	0.162	0.131	0.143	0.155
Panel C				**Credit: Formal Bank**				
ln(MA1996)	0.017	0.014	0.011	0.067	0.064	0.066*	0.063*	0.057*
	(0.01)	(0.01)	(7.07)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
First Stage F				18.49	11.14	26.95	12.36	41.70
				0.0000	0.0010	0.000	0.0000	0.000
R-squared	0.005	0.340		0.282	0.318	0.314	0.317	0.322
Control variables	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	212	212	212	212	212	212	212	212

Notes: This table reports the estimated effect of market access (MA1996) on technology adoption (Panel A), land sale (Panel B), and access to formal bank (Panel C). Columns (1) and (2) are estimated by OLS. Column (1) is a bivariate regression without any controls while column (2) includes state fixed effects and other controls. Column (3) is estimated by Oster's bias-adjusted OLS estimator. Columns (4) and (5) are estimated by 2SLS using railroad length and distance to GQ as IVs, respectively. Column (6) is estimated by 2SLS using predicted market access as the IV; which is estimated from a zero stage OLS regression including both railroad length and distance to GQ. Column (7) reports the Lasso estimates using a broad set of instruments that includes railroad length, distance to GQ, and each of their interactions with the exogenous control variables as IVs. The three IVs chosen by the lasso are: distance to GQ interacted with slope, distance to GQ interacted with elevation, and railroad length interacted with crop suitability. Column (8) is estimated by 2SLS using predicted market access estimated from a zero stage OLS regression including railroad length, distance to GQ, and all the lasso chosen IVs. The first stage Angrist-Pischke F test is reported. See notes from above tables for list of control variables. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

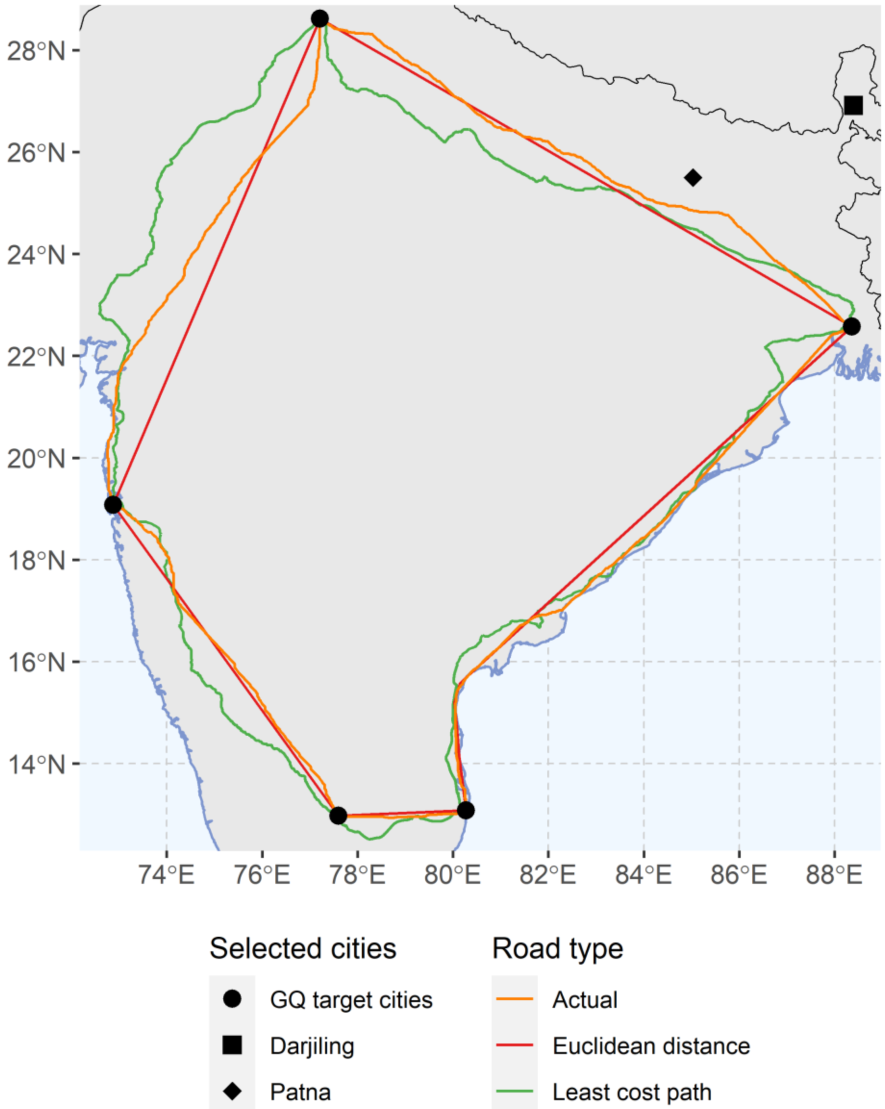
Figure 1

Kernel Density of Market Access in Districts with and without colonial railways



Source: Author calculation.

Figure 2. Golden Quadrilateral Road Networks



Source: Author elaboration using data from Ghani et al (2016).

NOT FOR PUBLICATION

Online Appendix

OA.1 Construction of the Market Access Index

Our Market Access measure is calculated as follows:

$$MA_{it} = \sum_{i \neq j} \left(\frac{1}{tt_{ijt}^{\theta}} \right) P_{jt}$$

where, MA_{it} is market access of District *i* at time *t*

tt_{ijt} is the travel time (in hours) between Districts *i* and *j* at time *t*

P_{jt} = population in destination District *j* at time *t*

$\theta =$ trade elasticity

To construct the Market Access measures, population and travel times were aligned as follows:

<table>
<tr><th>MA</th><th>Travel Time</th><th>Population</th></tr>
<tr><td>MA1996</td><td>1996</td><td rowspan="2">1991</td></tr>
<tr><td>MA2004</td><td>2004</td></tr>
<tr><td></td><td></td><td>2001</td></tr>
</table>

Note that for the main analysis, we focus on Market Access calculated using travel time from 1996 and population from 1991. We explore alternative travel times (2004) as robustness checks. Following Allen and Atkin (2016) we use trade elasticity equal to 1.5, and report results using trade elasticity equal to 3.8 as well.

OA.2 Calculation of the GQ-based instruments

Our identification strategy relies on two main sets of instrumental variables: one inspired by India’s Golden Quadrilateral (GQ) highways and the other by its colonial railways in the 1880s. Both are calculated using geospatial software. Here, we describe the calculation of the first instrumental variable.

For the first GQ IV, we construct a hypothetical linear network connecting the main cities targeted by the GQ project (see Figure 2). The target cities of the GQ include: New Delhi, Kolkata, Chennai, Mumbai and Bangalore. We construct two networks: one based on Euclidean distance and another “least cost path” derived from elevation and slope data. Our main results rely on the linear, Euclidean distance instrument but are robust to using the least-cost path instrument instead. It is also important to note that though the GQ highway project is a relatively recent investment, parts of it follows historical roads (Figure AF.1).

The least cost path is calculated based on a time cost raster method using the Cost Connectivity algorithm available in ESRI ArcGIS as the minimum time result from off road speeds on land. We construct the time cost raster by assigning a speed to cross each pixel based on Tobler's hiking function (1993) with a weight derived from historical land cover class (1900) following Ali et al. (2015). We use land cover from the HYDE model (Goldewijk et al. 2011) and we use slope data from Verdin (2007).

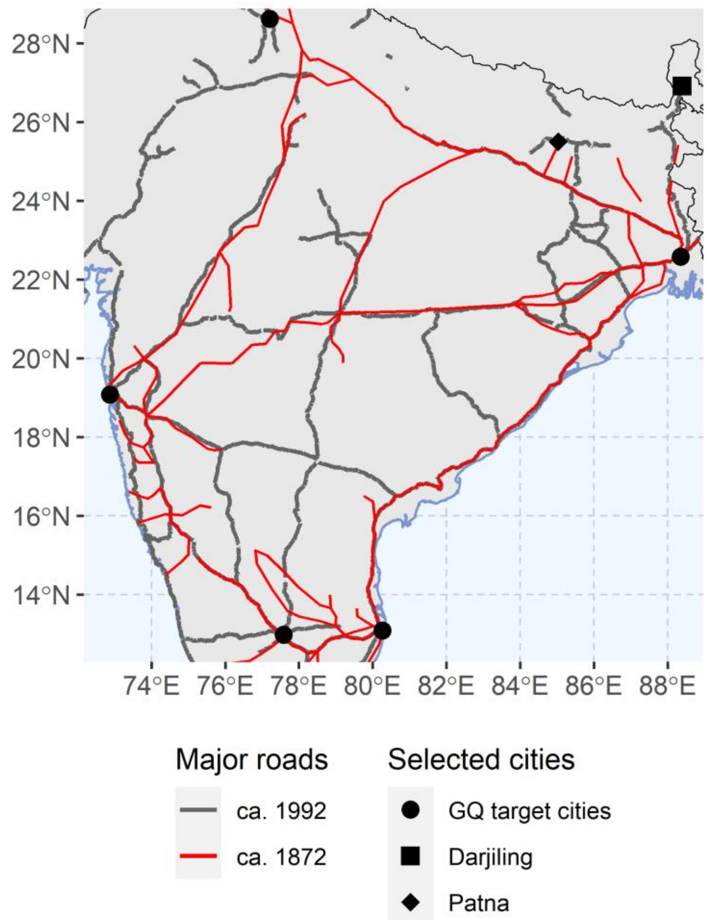
$$Hiking = 6 * e^{-3.5\ |s+0.05| * 0.6}$$

where s is mean slope.

We then calculate the distance between each district's centroid and each linear network. For robustness, we also consider the Euclidean distances from each district's centroids to the actual GQ highways, which are available from Ghani et al (2014).

Similar to Faber (2014), we address the concern of non-random local route placements on the way between targeted city nodes by constructing a hypothetical network based on Euclidean distance and least cost path based on elevation and slope. Faber uses a simple land cover model based on the engineering literature (see Jha et al., 2001; Jong and Schonfeld, 2003; cited in Faber 2014) and includes measures of slope, development, water and wetland, where the algorithm prefers short and flat routes.

Figure AF.1: Historical Roads of India, 1872 vs Golden Quadrilateral Roads in 1992



Note: Schwartzenberg roads in red are from 1872. Digital Chart of the World in black depict roads in 1992.
Source: Schwartzberg Atlas, Growth of Road Network, p. 125, available from the Digital South Asia Library. Digital Chart of the World.

Table A.1. Effects of Market Access on Land Inequality: OLS and Bias-Adjusted OLS Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.	**ln(Landgini 2012)**			**ln(Landless 2012)**		
Estimator	OLS	OLS	Bias-Adjusted	OLS	OLS	Bias-Adjusted
ln(MA1996)	**0.108*** **	**0.111*** **	**0.115*** **	**0.360*** **	**0.357*** **	**0.352*** **
	(0.02)	**(0.02)**	**(0.04)**	**(0.07)**	**(0.07)**	**(0.11)**
ln(Slope)		-0.037			-0.078	
		(0.03)			(0.09)	
Elevation		-0.138			0.148	
		(0.14)			(0.43)	
Crop Suitability		-0.002			-0.005	
		(0.00)			(0.00)	
ln(Rain)		0.039			0.153	
		(0.05)			(0.16)	
ln(Temperature)		-0.188			0.621	
		(0.35)			(0.96)	
ln(Rain CV)		0.089			0.256	
		(0.13)			(0.39)	
ln(Temp. seasonality)		-0.056			-0.064	
		(0.08)			(0.21)	
Pop. density, 1961		0.023***			0.037	
		(0.01)			(0.02)	
Constant	-1.942***	-0.980		-6.262***	-11.059*	
	(0.25)	(2.08)		(1.14)	(5.84)	
State Dummies	No	Yes	Yes	No	Yes	Yes
Observations	200	200	200	212	212	212
R-squared	0.166	0.380		0.167	0.428	

Notes: This table shows the estimated effects of market access on land inequality. Market Access is calculated using population in 1991 and travel time in 1996. Columns (1), (2) and (3) measure land inequality using a Gini index in 2012. Columns (4), (5), and (6) measure land inequality as the share of landless households in a district. The unit of analysis is the district level according to the 2001 Census boundaries. We control for geographic variables (slope, elevation, crop suitability, rain, temperature, rain coefficient of variation, temperature seasonality), population density in 1961, and state fixed effects. Columns (1), (2), (4) and (5) are estimated by OLS. Columns (3) and (6) are estimated by Oster's bias-adjusted OLS (Oster (2019)). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.2A. Effects of Market Access on Land Gini (IV Estimates)

	(1)	(2)	(3)	(4)	(5)
Dep. var.			**ln(Landgini 2012)**		
Estimator	2SLS	2SLS	2SLS	Lasso-IV	2SLS
ln(MA1996)	**0.244***	**0.232**	**0.241***	**0.255***	**0.197***
	(0.07)	**(0.09)**	**(0.06)**	**(0.06)**	**(0.05)**
ln(Slope)	-0.044	-0.043	-0.044	-0.044	-0.041
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Elevation	-0.191	-0.186	-0.190	-0.195	-0.172
	(0.15)	(0.15)	(0.15)	(0.16)	(0.14)
Crop Suitability	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(Rain)	0.053	0.052	0.053	0.054	0.048
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
ln(Temperature)	-0.428	-0.406	-0.421	-0.447	-0.343
	(0.42)	(0.44)	(0.41)	(0.42)	(0.37)
ln(Rain CV)	0.138	0.134	0.137	0.142	0.121
	(0.14)	(0.14)	(0.13)	(0.14)	(0.13)
ln(Temp. seasonality)	-0.024	-0.027	-0.025	-0.021	-0.035
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Pop. density, 1961	0.015**	0.016*	0.016**	0.015**	0.018***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-1.908	-1.806	-1.877	-2.271	-1.506
	(2.45)	(2.42)	(2.40)	(2.47)	(2.22)
First Stage					
ln(km of railroad)	0.117***		0.114***		0.192***
	(0.03)		(0.03)		(0.07)
ln(dist. to GQ)		-0.128***	-0.120**		-0.015
		(0.05)	(0.05)		(0.08)
ln(dist. to GQ) x ln(slope)				-0.131***	-0.112*
				(0.04)	(0.06)
ln(dist. to GQ) x Elevation				0.152	0.106
				(0.18)	(0.19)
ln(km of railroad) x Crop suitability				0.002***	-0.001
				(0.00)	(0.00)
Angrist-Pischke F	18.94	6.25	21.93	9.19	34.23
	0.0000	0.0134	0.0000	0.0000	0.0000
Observations	200	200	200	200	200
R-squared	0.239	0.263	0.247	0.216	0.321

Notes: Columns (1) and (2) are estimated by two-stage least squares (2SLS) using railroad length and distance to the Golden Quadrilateral as instrumental variables, respectively. Column (3) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero stage with both railroad length and distance to GQ. The Lasso in column (4) is estimated using a parsimonious set of instruments chosen by Lasso from a broad set that includes railroad length, distance to GQ, and each of their interactions with the exogenous control variables as instrumental variables. Lasso selects three instruments which are reported in the first stage in column (4). Column (5) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero-stage including railroad length, distance to GQ, and all the instruments chosen by Lasso. In columns (3) and (5), under the first stage, we report the zero stage estimated coefficients of the instrumental variables. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A. 2B. Effects of Market Access on Landlessness (IV Estimates)

	(1)	(2)	(3)	(4)	(5)
Dep. var.			ln(Landless 2012)		
Estimator	2SLS	2SLS	2SLS	Lasso-IV	2SLS
ln(MA1996)	**0.974*** **	**0.403***	**0.758*** **	**0.678*** **	**0.662*** **
	(0.23)	**(0.21)**	**(0.16)**	**(0.14)**	**(0.14)**
ln(Slope)	-0.079	-0.078	-0.078	-0.078	-0.078
	(0.11)	(0.09)	(0.10)	(0.10)	(0.10)
Elevation	-0.106	0.129	-0.017	0.016	0.022
	(0.50)	(0.41)	(0.44)	(0.42)	(0.42)
Crop Suitability	-0.010**	-0.005	-0.008**	-0.008**	-0.008**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(Rain)	0.120	0.151	0.132	0.136	0.137
	(0.19)	(0.15)	(0.17)	(0.16)	(0.16)
ln(Temperature)	-0.530	0.534	-0.127	0.021	0.051
	(1.34)	(0.99)	(1.12)	(1.06)	(1.03)
ln(Rain CV)	0.575	0.280	0.463	0.422	0.414
	(0.50)	(0.38)	(0.43)	(0.41)	(0.41)
ln(Temp. seasonality)	-0.104	-0.067	-0.090	-0.085	-0.084
	(0.26)	(0.19)	(0.22)	(0.21)	(0.21)
Pop. density, 1961	0.000	0.034	0.013	0.018	0.019
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	-14.063*	-10.750*	-12.808*	-13.660**	-12.255*
	(7.87)	(5.56)	(6.68)	(6.33)	(6.33)
First Stage					
ln(km of railroad)	0.114***		0.114***		0.186***
	(0.03)		(0.03)		(0.07)
ln(dist. to GQ)		-0.148***	-0.120**		-0.030
		(0.04)	(0.05)		(0.08)
ln(dist. to GQ) x ln(slope)				-0.131***	-0.106*
				(0.04)	(0.06)
ln(dist. to GQ) x Elevation				0.119	0.081
				(0.17)	(0.18)
ln(km of railroad) x Crop suitability				0.002***	-0.001
				(0.00)	(0.00)
Angrist-Pischke F	18.49	11.14	26.96	12.36	41.70
	(0.0000)	(0.0010)	(0.0000)	(0.0000)	(0.0000)
Observations	212	212	212	212	212
R-squared	0.154	0.427	0.312	0.3535	0.361

Notes: This table reports the estimated effect of market access on the land Gini index in 2012. Market Access is calculated using population in 1991 and travel time in 1996. We control for slope, elevation, crop suitability, rain, temperature, rain coefficient of variation, temperature seasonality, population density in 1961, and state fixed effects. Columns (1) and (2) are estimated by two-stage least squares (2SLS) using railroad length in the 1880s in a district, and distance of a district to the nearest arm of Golden Quadrilateral as instrumental variables, respectively. Column (3) is estimated using predicted Market Access as an instrumental variable from a zero stage with both railroad length and distance to GQ. The Lasso in column (4) is estimated using the instruments chosen by Lasso from a broad set that includes railroad length, distance to GQ, and each of their interactions with the exogenous control variables as instrumental variables. Lasso selects three instruments which are reported in the first stage in column (4). Column (5) is estimated using predicted Market Access as an instrumental variable. The predicted Market Access are fitted values from a zero-stage including railroad length, distance to GQ, and all the instruments chosen by Lasso. In columns (3) and (5), under the first stage, we report the zero stage estimated coefficients of the instrumental variables. The first stage reports the estimated coefficients of the instrumental variables from a regression of market access on the controls and instruments. T Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Estimates Using Travel Times in 2004 : Lasso IV

	(1) ln(Landgini, 2012)	(2) ln(Landless, 2012)	(3) ln(Tech use, 2012)	(4) ln(Land Sale, 2012)
ln(MA2004)	**0.277***	**0.712***	**0.382***	**0.879**
	(0.07)	**(0.17)**	**(0.09)**	**(2.35)**
ln(Slope)	-0.063**	-0.120	-0.097*	-0.656
	(0.03)	(0.10)	(0.05)	(1.12)
Elevation	-0.170	0.054	0.272	3.612
	(0.15)	(0.41)	(0.24)	(5.28)
Crop Suitability	-0.003**	-0.008**	-0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.05)
ln(Rain)	0.086	0.186	0.034	-1.114
	(0.06)	(0.17)	(0.10)	(2.14)
ln(Temperature)	-0.491	-0.147	0.526	1.586
	(0.44)	(1.08)	(0.66)	(14.85)
ln(Rain CV)	0.143	0.451	0.445*	10.237*
	(0.15)	(0.42)	(0.24)	(5.65)
ln(Temp. seasonality)	0.014	-0.036	-0.074	-4.464
	(0.08)	(0.22)	(0.11)	(3.25)
Pop. density, 1961	0.017***	0.026	-0.006	-0.088
	(0.01)	(0.02)	(0.01)	(0.31)
Constant	-2.983	-14.393**	-11.56***	-16.911
	(2.65)	(6.53)	(3.77)	(90.32)
First Stage				
ln(dist. to GQ) x ln(Slope)	-0.132***	-0.127	-0.127	-0.127
	(0.04)	(0.04)	(0.04)	(0.04)
ln(dist. to GQ) x Elevation	0.162	0.111	0.111	0.111
	(0.17)	(0.17)	(0.17)	(0.17)
ln(km of railroad) x Crop Suitability	0.001***	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Angrist-Pischke F	7.68	9.93	9.93	9.93
	0.0001	0.0000	0.0000	0.0000
Observations	200	212	212	212
R-squared	0.181	0.331	0.524	0.151

Note: All columns are estimated by Lasso and include state fixed effects. Weak identification test (Angrist-Pischke F) and R-squared are from 2SLS using the IVs chosen by Lasso. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Effects of Market Integration: Market Access Measure Based on trade elasticity 3.8

	(1)	(2)	(3)	(4)
	ln(Landgini, 2012)	ln(Landless, 2012)	ln(Tech use, 2012)	ln(Land Sale, 2012)
ln(MA1996), θ = 3.8	**0.092****	**0.306*****	**0.182*****	**0.775**
	(0.04)	**(0.11)**	**(0.07)**	**(1.07)**
ln(Slope)	-0.003	0.079	0.019	-0.210
	(0.04)	(0.14)	(0.08)	(1.25)
Elevation	-0.560**	-1.117	-0.438	0.339
	(0.27)	(0.75)	(0.42)	(7.08)
Crop Suitability	-0.003*	-0.008*	-0.002	-0.009
	(0.00)	(0.00)	(0.00)	(0.05)
ln(Rain)	-0.099	-0.303	-0.257*	-2.334
	(0.07)	(0.24)	(0.15)	(2.73)
ln(Temperature)	-1.452**	-3.052	-1.285	-7.621
	(0.71)	(1.94)	(1.12)	(19.96)
ln(Rain CV)	0.403*	1.146*	0.880**	12.487**
	(0.21)	(0.64)	(0.38)	(6.34)
ln(Temp. seasonality)	-0.133	-0.357	-0.265	-5.270
	(0.09)	(0.29)	(0.17)	(3.37)
Pop. density, 1961	0.006	-0.018	-0.034	-0.241
	(0.01)	(0.04)	(0.02)	(0.38)
Constant	7.254*	13.753	5.086	40.458
	(3.78)	(10.45)	(5.94)	(113.36)
First Stage				
Angrist-Pischke F	9.06	10.94	10.94	10.94
	0.003	0.0011	0.0011	0.0011
Observations	200	212	212	212
R-squared	-0.126	-0.147	0.161	0.165

Notes: Estimated by 2SLS using predict market access as the instrumental variable. Predicted market access is estimated from a zero-stage including railway length, distance to GQ, and the IVs chosen by Lasso (distance to GQ x Pop. density, 1961; and railway length x Crop suitability). All columns are estimated by 2SLS and include state fixed effects. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Estimates Controlling for the Colonial Land Revenue System

	(1)	(2)	(3)	(4)
	ln(Landgini, 2012)	ln(Landless, 2012)	ln(Tech use, 2012)	ln(Land Sale, 2012)
ln(MA1996)	**0.278*** **	**0.386**	**0.285** **	**-0.403**
	(0.06)	**(0.24)**	**(0.14)**	**(2.31)**
British control (yes=1)	-0.061	0.075	-0.020	2.926*
	(0.04)	(0.13)	(0.08)	(1.60)
No landlord (yes=1)	-0.084	-0.019	0.074	3.497
	(0.05)	(0.15)	(0.10)	(2.15)
ln(Slope)	-0.030	-0.078	-0.084*	-1.226
	(0.03)	(0.09)	(0.05)	(1.21)
Elevation	-0.242	0.169	0.283	5.813
	(0.16)	(0.42)	(0.25)	(5.52)
Crop Suitability	-0.003**	-0.005	-0.001	-0.001
	(0.00)	(0.00)	(0.00)	(0.05)
ln(Rain)	0.039	0.146	0.028	-0.383
	(0.06)	(0.15)	(0.09)	(2.21)
ln(Temperature)	-0.542	0.554	0.816	6.563
	(0.43)	(1.04)	(0.67)	(15.37)
ln(Rain CV)	0.129	0.340	0.340	10.346*
	(0.15)	(0.38)	(0.23)	(5.61)
ln(Temp. seasonality)	-0.042	-0.045	-0.090	-3.010
	(0.08)	(0.21)	(0.12)	(3.38)
Pop. density, 1961	0.011	0.038	-0.006	0.098
	(0.01)	(0.03)	(0.02)	(0.35)
Constant	-1.657	-11.039**	-10.019***	-46.566
	(2.54)	(5.51)	(3.74)	(92.78)
First Stage				
ln(dist. to GQ) x ln(Slope)	-0.129***	-0.128***	-0.128***	-0.128***
	(0.04)	(0.03)	(0.03)	(0.03)
ln(dist. to GQ) x Elevation	0.217	0.161	0.161	0.161
	(0.18)	(0.17)	(0.17)	(0.17)
ln(dist. to GQ) x ln(Slope)	0.002***	0.002***	0.002***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)
Angrist-Pischke F	8.05	10.96	10.96	10.96
	0.0000	0.0000	0.0000	0.0000
Observations	200	212	212	212
R-squared	0.1756	0.3458	0.5119	0.1446

Note: All columns are estimated by Lasso and include state fixed effects. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: Including population density from 2011 and proportion of Muslims as additional controls

	(1)	(2)	(3)	(4)
Dep. var.	ln(Landgini, 2012)	ln(Landless, 2012)	ln(Tech use, 2012)	ln(Land Sale, 2012)
ln(MA1996)	**0.273*** **	**0.691*** **	**0.327** **	**-0.162**
	(0.06)	**(0.15)**	**(0.13)**	**(2.37)**
ln(Slope)	-0.012	-0.003	-0.048	-1.318
	(0.03)	(0.09)	(0.05)	(1.24)
Elevation	-0.217	-0.009	0.267	4.322
	(0.15)	(0.40)	(0.24)	(5.27)
Crop Suitability	-0.003**	-0.007**	-0.001	-0.003
	(0.00)	(0.00)	(0.00)	(0.05)
ln(Rain)	0.042	0.160	0.036	-0.874
	(0.06)	(0.16)	(0.09)	(2.21)
ln(Temperature)	-0.478	0.083	0.774	4.327
	(0.44)	(1.04)	(0.69)	(15.09)
ln(Rain CV)	0.166	0.371	0.333	8.771
	(0.14)	(0.41)	(0.25)	(5.80)
ln(Temp. seasonality)	-0.012	-0.015	-0.073	-5.132
	(0.08)	(0.20)	(0.11)	(3.24)
Pop. density, 1961	0.335***	0.716*	0.145	-9.044*
	(0.12)	(0.37)	(0.25)	(4.78)
Pop. density, 2011	-0.172***	-0.373*	-0.081	4.828*
	(0.07)	(0.20)	(0.13)	(2.59)
Muslim prop.	0.251***	0.989***	0.529***	-4.683*
	(0.10)	(0.25)	(0.14)	(2.82)
Constant	-2.566	-14.956**	-10.655***	-3.339
	(2.53)	(6.11)	(3.76)	(89.29)
First Stage				
ln(dist. to GQ) x ln(Slope)	-0.101***	-0.102***	-0.102***	-0.102***
	(0.04)	(0.03)	(0.03)	(0.03)
ln(dist. to GQ) x Elevation	0.067	0.049	0.049	0.049
	(0.17)	(0.16)	(0.16)	(0.16)
ln(dist. to GQ) x ln(Slope)	0.002***	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Angrist-Pischke F	8.02	10.66	10.66	10.66
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	200	212	212	212
R-squared	0.2310	0.4024	0.5478	0.1567

Note: All columns are estimated by 2SLS and include state fixed effects. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.7 : Using Distance to Least-Cost GQ Network as Instrument

	(1)	(3)	(5)	(7)
Dep. var.	ln(Landgini, 2012)	ln(Landless, 2012)	ln(Tech use, 2012)	ln(Land Sale, 2012)
ln(MA1996)	**0.188*** **	**0.653*** **	**0.334*** **	**1.983**
	(0.06)	**(0.17)**	**(0.10)**	**(2.05)**
ln(Slope)	-0.041	-0.078	-0.074	-0.607
	(0.03)	(0.10)	(0.05)	(1.12)
Elevation	-0.169	0.026	0.264	3.094
	(0.14)	(0.42)	(0.25)	(5.19)
Crop Suitability	-0.003**	-0.008**	-0.001	-0.010
	(0.00)	(0.00)	(0.00)	(0.05)
ln(Rain)	0.047	0.137	0.008	-1.237
	(0.05)	(0.16)	(0.09)	(2.15)
ln(Temperature)	-0.326	0.069	0.672	-0.343
	(0.37)	(1.03)	(0.64)	(14.30)
ln(Rain CV)	0.117	0.409	0.414*	10.793*
	(0.13)	(0.40)	(0.23)	(5.60)
ln(Temp. seasonality)	-0.037	-0.083	-0.098	-4.600
	(0.07)	(0.21)	(0.12)	(3.21)
Pop. density, 1961	0.019***	0.019	-0.009	-0.165
	(0.01)	(0.02)	(0.01)	(0.33)
Constant	-1.426	-12.200*	-10.039**	-27.137
	(2.18)	(6.32)	(3.95)	(87.85)
First Stage				
Angrist-Pischke F	24.9	28.46	28.46	28.46
	0.0000	0.0000	0.0000	0.0000
Observations	200	212	212	212
R-squared	0.333	0.365	0.524	0.162

Estimated by 2SLS using predict market access as the instrumental variable. Predicted market access is estimated from a zero stage including railway length, distance to least cost network, and the IVs chosen by Lasso (distance to GQ x Pop. density, 1961; and railway length x Crop suitability). All columns are estimated by 2SLS and include state fixed effects. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Summary Statistics

	Mean	Std. Dev.	Min	Max
Outcome Variables				
Land Gini (index), 2012	0.78	0.13	0.46	0.99
Landlessness (%), 2012	56.36	24.71	4.29	98.63
Technology use (%), 2012	73.19	21.19	9.09	100.00
Land sale (rupees), 2012	495,037	1,427,056	0	12,900,000
Treatment				
Market access (index), 1996	6,862,575	4,261,530	817,823	22,800,000
Controls				
Mean Slope (%)	5.74	7.87	0.97	53.20
Elevation (1,000 meters)	0.35	0.35	0.01	2.94
Crop suitability index	63	19	1	100
Rainfall (millimeters)	1,151	678	209	4,157
Rainfall coefficient of variation	254	24	80	289
Temperature (Celcius)	120	21	57	156
Temperature seasonality (st. dev. Celcius)	4421	1687	926	7399
Population density, 1961	0.33	0.71	0.02	9.50
Number of observations	200			