What Explains the Geographic Variation in Corporate Investment?*

Shohini Kundu[†] Nishant Vats[‡]

Abstract

We show that history can explain the geographic concentration of investment over and above traditional agglomerative forces, geography, and expectations. We use spatial variation in direct and indirect British rule in India to identify differences in historical circumstances. Using this within-country variation in historical circumstances, combined with a local identification approach and instrumental variable strategy, we explain the spatial differences in investment. Differences in historical origins can explain 13% of total geographic variation in investment. Moreover, investment is 8-10% lower in direct ruled areas. Our results indicate that history can have long-run consequences through its effect on economic organizations.

*We thank Rainald Borck, Francesco D'Acunto, Zhiguo He, Lakshmi Iyer, Ralph Koijen, Yueran Ma, Nagpurnanand Prabhala, Raghuram Rajan, James Robinson, Amir Sufi, Nico Voigtländer, Michael Weber, Xiao Zhao, and Eric Zwick for helpful comments and suggestions. We thank seminar participants at the 100 Years of Economic Development Conference, Webinar series in Finance and Development (WEFIDEV), Asian Bureau of Finance and Economic Research (ABFER) Conference, 2021 Transatlantic Doctoral Conference at the London School of Business, 10th European Meeting of the Urban Economic Association, the NSE-NYU Conference on Indian Financial Markets, 2020 and the 12th Emerging Markets Conference. This research was funded in part by the John and Serena Liew Fellowship Fund at the Fama-Miller Center for Research in Finance, University of Chicago Booth School of Business, the Wheeler Institute for Business and Development at the London Business School and the NSE-NYU Stern Initiative on the Study of Indian Financial Markets. We do not have any conflicts of interest to disclose. We take responsibility for all errors.

[†]Anderson School of Management, University of California, Los Angeles. email: shohini.kundu@anderson.ucla.edu

[‡]Olin Business School, Washington University - St. Louis, United States. email: vats@wustl.edu

1 Introduction

What explains geographic concentration in investment? What role does history play? It has been established that investment is geographically concentrated. From Steel City to Silicon Valley, the formation of economic clusters has garnered attention from academics and policymakers alike. Understanding the forces behind the development of such clusters is fundamental to understanding the micro-foundations of geographic inequality – why are certain areas richer than others, and what can be done to address these differences? Thus far, the literature has focused extensively on the role of agglomerative forces – the availability and cost of factors of production – and geographic advantages to explain the rise and fall of such clusters. In this paper, we diverge from the extant literature, focusing on a novel aspect that can explain the geographic concentration of investment: history. We consider how the eventual investment concentration depends on the initial starting point, set by historical factors. Specifically, we emphasize the role of the destruction of well-established economic organizations, resulting from differences in historical circumstances, as key factors in explaining the investment concentration today.

The central role of history may seem obvious if investment is assumed to follow a path-dependent process. However, establishing the empirical relationship between history and investment is difficult. A key element in the theoretical models of firms' choice of location is the existence of multiple equilibrium. Krugman (1991) argues that the eventual choice of the equilibrium can either be driven by the history or self-fulfilling expectations. Hence, the empirical challenge lies in disentangling the two forces. Moreover, to clearly establish the relevance of history, one needs to rule out the effect of confounding unobservables such as agglomerative forces and geographic advantage.

In this paper, we attempt to address this issue and show that historical circumstances can explain geographic concentration over and above the traditional agglomerative forces, geographic advantages and self-fulfilling expectations. We do this by using a two-pronged approach. First, we combine within country analysis with plausibly exogenous variation in historical circumstances, originating from direct and indirect colonial rule. Second, in addition to using conventional approaches of comparison, we use (1) a local identification approach of comparing neighboring areas to rule out concerns related to differences in agglomerative forces, geography and expectations, and

(2) an instrumental variable strategy to rule out concerns related to selection bias and simultaneity. Hence, our empirical strategy is better adept at solving the difficult identification problems associated with establishing a direction of causality between history and economic activity, in general, and, in particular, investment concentration.

This paper uses the within-country geographic variation in historical circumstances to explain spatial differences in investment. Colonial occupation of India provides such an environment. Before the onset of the British occupation of India starting in 1757, the Indian subcontinent was governed by local rulers. During the colonial era, different areas of the Indian subcontinent fell under the "direct rule" of the British, or, "indirect rule," under the administration of native rulers. All areas, regardless of their colonial history, were brought under an identical legal and administrative framework after independence in 1947. Moreover, we verify that the direct and indirect ruled areas were similar across several observable dimensions before the onset of colonial rule. Therefore, India provides an ideal laboratory to examine the consequences of differences in historical circumstances on geographic variation in investment concentration in the present.

We begin with an aggregate analysis, showing that investment is concentrated within Indian states. Using data on district-level corporate investment, we compute a state-level measure of investment concentration, using the Herfindahl-Hirschmann Index (HHI). We compare the investment HHI with three benchmark measures – (1) equal investment in all districts, $\frac{1}{N}$, (2) investment proportional to geographical area of district relative to the state, and (3) investment proportional to the population of a district relative to the state. We conclude that investment is geographically concentrated within states, relative to a frictionless spatial equilibrium. Moreover, we show that states with a larger proportion of districts historically under direct British rule exhibit a higher geographic concentration of investment. Specifically, the level of the within-state investment concentration is 20 percentage points higher relative to any of the benchmarks. Moreover, the proportion of districts under direct rule can explain 13% of total variation in within-state investment concentration.

States with a greater proportion of direct ruled districts exhibit greater investment concentration. We argue that greater investment concentration in states with a larger proportion of direct ruled districts is driven by higher investment in indirect ruled districts. A direct comparison of direct and indirect ruled districts combined with project level data indicates that the size of investment projects in direct ruled areas is 8.8% lower than the size of investment projects in indirect ruled areas within a state. This estimate is robust to the inclusion of firm $\times$ year and state $\times$ year fixed effects. Moreover, we characterize that the differences in investment occur at both the intensive and the extensive margins. On the extensive margin, we find that projects are 25% less likely to be announced in direct ruled districts relative to indirect ruled districts, within a state. On average, total investment is ₹ 16 billion ($ 0.2 billion) lower in direct ruled districts relative to indirect ruled districts within the same state.

We verify that our baseline findings are unlikely to be driven by geographic differences across districts, differences in law, differences in the distances to the state capitals, and spatial trends. Moreover, we conduct a placebo test, in which we run 10,000 simulations randomizing whether a district was under direct rule or not. We are able to generate an effect of a size at least as large as the baseline estimate only in 0.1% of cases, thereby ruling out issues related to the spuriousness and spatial autocorrelation of investment. The Oster (2019) test for our baseline analysis raises our confidence that the baseline results are unlikely to be driven by omitted variables. Moreover, we verify that the direct and indirect ruled areas are similar across several observable characteristics, such as geography, distance from the state capital, the religion of precolonial rulers, and the colonial population composition.

While these tests are informative and raise confidence in our results, we cannot completely rule out differences among districts along all observable and unobservable dimensions. Our baseline analysis assumes that one district in India is as good a control as any other district within a state. A direct comparison of direct and indirect districts may potentially bias our inference in the presence of systematic differences between direct and indirect ruled districts. Hence, we address concerns of selection and omitted variables by focusing on contiguous direct-indirect ruled district pairs, separated only by administrative borders within a state. We compare a firm's investment in direct and indirect ruled districts within a district-pair. Specifically, we include firm $\times$ district-pair $\times$ year fixed effect, allowing us to identify the estimate using variation in the size of investment projects announced by the same firm within a district-pair. Such an approach allows us to implicitly control for traditional agglomerative forces of Marshall (1920), geographic features, investment opportunities, and expectations that are likely to be similar across contiguous district-pairs. Moreover, whether a district within a contiguous direct-indirect ruled pair was under direct rule during the colonial period is likely to be a matter of chance. Hence, indirect ruled districts are a valid counterfactual to the contiguous direct ruled districts. Our local identification approach suggests that the projects announced in direct ruled districts are 10.8% smaller in size relative to the projects announced in indirect ruled districts by the same firm within a contiguous district-pair.

The comparison of direct ruled districts with indirect ruled districts may still be prone to selection bias, hindering our ability to interpret the baseline effect as causal. While the low value of the Moran I statistic indicates that selection is likely to be of little concern, it may be unreasonable to assume that the direct control by British in India was random. We address concerns of selection through an instrumental variable strategy.

We exploit a unique feature of British annexation policy in India, the Doctrine of Lapse. Between 1848 and 1856, the Doctrine of Lapse allowed the governor-general of British India to annex Indian princely states where the ruler died without a natural heir. The relevance condition posits that the death of a ruler without a natural heir is associated with a territory coming under direct British rule. We verify the relevance condition associated with this instrument in the first stage. Our two stage least square estimates (2SLS) indicate that investment is lower in direct ruled districts relative to indirect ruled districts within a state, validating our interpretation of the relation between contemporary corporate investment and direct British rule being causal. An important assumption of this test is the exclusion restriction, which posits that the death of a ruler without a natural heir affects current corporate investment only through the territory being under direct British rule. We verify this assumption through two falsification tests. In the first falsification test, we directly study the relation between investment and the death of the ruler without a natural heir in indirect ruled districts in periods when the policy was not applicable. In the second falsification test, we directly study the relation between investment and the death of the ruler with a natural heir in indirect ruled districts when the Doctrine of Lapse was applicable. Both tests yield null results, supporting the exclusion assumption. The first test addresses the concern that the death of the ruler is attributable to other features, such as environmental factors or genetics, which may directly affect postcolonial investment. The second test addresses the concern that the death of the ruler between 1848 and 1856, regardless of the presence of a natural heir, may reflect geography-specific conditions during that period which can affect long-term outcomes through channels other than direct rule.

We explore the underlying mechanism through which direct British rule affects investment in the present. We introduce a new channel through which history can explain the geographic concentration of investment: the destruction of well-established economic organizations. We rule out alternative mechanisms that may explain our results including the geographic differences in law and enforcement, provision of public goods, and trust.

We show that the East India Company (EIC) consolidated economic power through the annexation of cotton producing territories. Specifically, a cotton producing district was 40% more likely to be under direct British rule, relative to indirect rule. This allowed the British to directly control the supply of cotton, securing a monopoly on the supply of Indian goods and products (Sahoo, 2015). The direct control of cotton producing territories allowed the British to meet their objectives of protecting the interests of the British textile industry and increase Britain's share of global trade. Specifically, we argue that areas under direct colonial rule were subject to economic policies and practices that dismantled the well-established local economic organizations centered around the cotton industry. The destruction of flourishing and dominant economic industries resulted in significant economic losses, borne by the native population.

We use the invention of the cotton gin in 1794 as a natural experiment to study how the incentives of the British to expropriate the Indian cotton textile industry changed with the widespread adoption of the cotton gin in the US. As the cotton farmed in India was of the long-staple variety, it did not benefit from this invention. Therefore, we hypothesize that the British had weaker incentives to expropriate the Indian cotton textile sector after the invention of the cotton gin that made the US the key supplier of raw cotton. Combining the timing of the widespread adoption of the cotton gin by 1800, with historical data on Indian domestic exports of agricultural commodities from 1884 until 1920, we document that the domestic cotton exports of provinces annexed before 1800 are lower than the domestic exports of provinces annexed after 1800. This finding is consistent with our hypothesis. We also note that this effect is not driven by a lower likelihood of cotton-producing areas falling under direct British rule after 1800.

We show that these economic losses have persisted to the present. Using precolonial cotton production as an instrument for direct British rule, we show that a firm reduces its project size by 20% in direct ruled areas relative to indirect ruled areas, within a pair of adjacent districts in the present. Further, we examine the impact on present day corporate investment among direct ruled areas that were annexed before 1800 relative to areas annexed after 1800, to study whether the reduced investment in direct ruled areas is primarily driven by cotton producing areas that were annexed before 1800, when the British may have had greater incentives to destroy the existing cotton industry. Indeed, we find that the present day lower investment in historically direct ruled areas is driven by cotton producing districts that were annexed before 1800. We argue that the economic losses have endured to the present as the destruction of strong economic organizations (1) hampers the intergenerational transfer of skills and knowledge, and (2) disrupts the natural Marshallian process that develops over time and explains the agglomeration and dominance of industries.

Related Literature: This paper is primarily related to two strands of the literature: geographic concentration of economic activity and the long-run effects of colonial rule.

First, we contribute to the literature examining the determinants of geographic concentration of economic activity by introducing a novel aspect that can explain the geographic concentration of investment: historical circumstances. The extant literature since Marshall (1920) has focused on the role of the cost and availability of the factors of production and geographic advantages as the primary determinant of emergence of clusters (Ellison and Glaeser, 1997; Duranton and Overman, 2005; Ellison, Glaeser and Kerr, 2010). We argue that the historical episodes of direct colonial rule can determine the concentration of investment over and above these traditional agglomerative forces. The results support the theoretical models of Ethier (1982), Panagariya (1986), Arthur (1986), and Krugman (1987), which highlight the role of history in the eventual choice of an economy's equilibrium. Our local identification approach, using firm $\times$ district-pair $\times$ year fixed effects, allows us to disentangle the importance of differences in historical circumstances from differences in traditional agglomerative forces of Marshall (1920), and, differences in geographic features. Moreover, our local identification approach allows us to distinguish between the history versus expectations effect discussed in Krugman (1991), as we compare adjacent districts that are likely to have same similar investment opportunities, hence, have similar expectations, but differ in their historical exposure to direct British rule. Therefore, the key contribution of this paper is to highlight the importance of geography as an important determinant of geographic concentration.

Second, this paper is related to the literature examining the long-run effects of the presence of a large foreign enterprise. Specifically, we show that foreign enterprises with high market power can have negative long-run consequences on local investment. This occurs through the dismantling of pre-existing and competing economic organizations. This result is in contrast with prior work which shows that foreign enterprises with high market power can have positive long-run effects on local economic development in the presence of high labor mobility (Méndez-Chacón and Van Patten, 2019), or when they organize colonial economic activity to maximize their economic returns (Dell and Olken, 2020). Our primary result is consistent with other works (Nunn, 2008; Dell, 2010; Bobonis and Morrow, 2014; Lowes and Montero, 2020)), which document negative, persistent development effects of exploitative foreign enterprises. These papers focus on the coercion of labor – the threat or use of force to compel workers to enter into an employment relationship – as the key driver of trust which explains persistent detrimental effects. In contrast, using the case of the cotton industry, we show that the long-run detrimental effects of foreign enterprises can originate from the dismantling of pre-existing economic organizations, which are likely to pose a threat to the monopoly power of foreign enterprises. We argue that the long-run detrimental effects of the destruction of well-established economic organizations originate from the destruction of stable human and financial capital stock.

Our channel – destruction of well-established economic organizations – supplement the longstanding literature, exploring the long-run effects of colonial rule by identifying the mechanism through which colonial regimes can affect contemporary economic outcomes. This adds to the several other mechanisms through which colonial regimes persist such as mistrust (Nunn and Wantchekon, 2011) economic and physical infrastructure (Donaldson, 2018; Dell and Olken, 2020; Berger, 2009), health and education infrastructure (Huillery, 2009), legal regimes and property protection (Acemoglu, Johnson and Robinson, 2001; Porta et al., 1998; Banerjee and Iyer, 2005).

Our paper is closest to Iyer (2010), which examines the long-run consequences of direct colonial rule in India. However, we highlight a new channel not discussed in Iyer (2010) – colonial destruction of well-established industries that can explain long-run consequences of direct rule. Using data on public goods provision across Indian districts from 1961 until 1991, Iyer (2010) documents differences in public goods provision across direct and indirect ruled districts. These differences diminish substantially over time, exhibiting convergence, consistent with policies of equalization followed by postcolonial governments (Iyer, 2010; Banerjee and Somanathan, 2007). Hence, such differences in public goods provision are muted for our sample period – which we verify – and are unlikely to explain the persistent differences in investment concentration, even after more than two decades since 1991.

This paper is organized as follows. Section 2 provides a short description of the historical setting. Section 3 describes the datasets that are utilized in this project. Section 4 presents our key results. Section 5 presents the evidence from bordering districts. Section 6 presents the results from the instrumental variable regressions. Section 7 discusses the mechanism. Section 8 concludes.

2 Historical Setting

In this section, we provide a short description of the setting for the paper. We chronicle British annexation of India and expound on the process of colonialism in Appendix A. The Indian subcontinent was initially governed by local rulers. During the colonial era, districts came either under indirect rule or direct rule. We argue that direct and indirect ruled districts differ on two dimensions. First, the direct ruled districts faced greater economic exploitation. Second, districts that fell under indirect rule, where precolonial rulers retained administrative powers, had greater state capacity, while districts that fell under direct rule of the British had lower state capacity (Iyer, 2010).

2.1 An Overview of India under British Rule (1757-1947)

British colonial rule over the Indian subcontinent began in 1757 and lasted until 1947. After the decline of the Mughals and the Maratha Empire, India was divided into several small states ruled by royal families, referred to as *princely states*. The British Empire consolidated its grip on political control across present day India, Bangladesh, Burma, and Pakistan in "British India" under three broad policies of annexation, namely, the Subsidiary Alliance, the Doctrine of Lapse, and Annexation under Misrule. With these policies, certain areas that had been annexed by the British were integrated under direct rule, while remaining areas, were granted indirect rule, in which authority was delegated to local princes (princely states). Figure 1 presents the geographic distribution of direct and indirect ruled areas.

The three waves of annexation are characterized by three different policies. Under Subsidiary Alliance, the East India Company entered contracts with princely states, providing the latter with the subsidiary militia for protection and necessary defense for payment. In the event of default by the Indian ruler, a part of their territory was surrendered to the British. The second wave of British annexation, *subordinate isolation*, began in 1818, lasting until the Indian Mutiny of 1857. This phase was marked by the policy of Doctrine of Lapse between 1848 and 1856. Under the Doctrine of Lapse, any subordinate princely state would automatically lapse to the East India Company if the ruler died without a natural heir. Power was transferred from the East India Company to the British Crown after the Sepoy Mutiny of 1857 putting a formal end to active British annexation of India.[1] Despite the Queen's proclamation of 1858 which ruled out any future annexation after the Sepoy Mutiny, the Governor-General did retain the power to interfere in the internal matters of the princely states in case of a "misrule." Ashton (1977) documents active interference by the British colonial power in the internal affairs of princely states, e.g., Lord Curzon, the Governor-General of India from 1899 to 1905 forced fifteen rulers to abdicate, during his tenure on account of misrule.

Princely states were heterogeneous – some consisted of a few villages, while other kingdoms presided over thousands of square miles. Princely states had varying degrees of legal autonomy, ranging from first-class, wherein the state could try criminal cases to third-class states whereby only small civil cases could be adjudicated by the ruler. Iyer (2010) notes that the princely states constituted approximately 45% of the total geographic area of present day India, and 23% of total population in 1911. Additionally, she notes the presence of princely states across India with a high concentration in central and western India. The Foreign Office recognized about 680 Indian princely states in the year 1910 (Iyer, 2010).

After Indian Independence in 1947, areas under direct British rule federated. The princely states, however, could choose to join the Indian Union or remain independent. Prime Minister Jawaharlal Nehru and the Home Minister Sardar Vallabhbhai Patel integrated all princely states into the Indian Union by 1950. Thereafter, princely rulers no longer maintained sovereignty, and the indirect ruled districts faced similar political and legal structures as the direct ruled districts within modern-day states.

2.2 Colonial Differences

The differences between direct and indirect British rule originate from several sources, both political and economic.

Politically, rulers of princely states were under constant threat of being annexed in case of a misrule. *Annexation on Misrule* incentivized princely states to provide better governance and

[1] This is referred as the Government of India Act 1858 – an Act of the British Parliament, which nationalized the East India Company and granted the British monarch supreme authority over India, as well as power and possessions of the East India Company.

institutions. Princely states had between four to five rulers during 1858-1947, whereas states under direct British rule were governed by 24 Governor-Generals during the same period (Iyer, 2010). The longer tenure of native rulers granted them greater incentives to plan and invest for long-term development. Hence, the combination of a longer tenure, as well as the threat of annexation contributed to development of better early institutions and governance policy in indirect ruled areas, resulting in greater state efficiency. Iyer (2010) documents persistence of the differences in public goods provision between direct and indirect ruled districts post independence. However, she notes that these differences diminish substantially over time, exhibiting convergence. Using data from 1971 to 1991, Banerjee and Somanathan (2007) also document convergence in public goods provision across regions, consistent with policies of equalization followed by postcolonial governments.[2] We verify that the difference in public goods provision are muted for our sample period.

Economically, areas under direct colonial rule faced the brunt of protectionist policies, restrictive contracts, and exploitative practices, causing the native population to experience significant economic losses. The native population in areas under direct rule were also more likely to face physical abuse, affecting their marginal productivity of labor. Additionally, the British neither encouraged the industralization of indigenous industries in India, nor developed institutions to promote them. These factors are discussed in the context of cotton production in section **??**, although the mechanism applies in other contexts as well.[3]

2.3 What Determines Direct Rule?

In this section, we explore if direct rule can be explained by a multitude of factors, such as geographic features, distance to state capital, religion of the precolonial ruler, and population distribution during colonial times. Table 1 presents results from a linear probability model, regressing the likelihood of an area being under direct rule on these characteristics. Table 1 indicates that the majority of these characteristics cannot predict direct rule except Muslim ruler. The likelihood of a territory being under direct rule increases if the area was ruled by a Muslim ruler. With the exception of the

[2]Specifically, Banerjee and Somanathan (2007) document that the two decades after 1971 there was a dramatic expansion in rural infrastructure. By 1991 primary schools were available in nearly 74% of all villages. Over the two decades, the share of villages with electricity connections increased from 18% to 70% and there was a nine-fold increase in access to piped water.

[3]The EIC became the "de facto ruler of Bengal, where it established a monopoly in grain trading and prohibited local traders and dealers from 'hoarding' rice'" (Zingales (2017)). The combination of high taxation and monopolistic power contributed to the Great Bengal Famine, during which one out of three Bengalis died of starvation.

religion of the precolonial ruler, the direct and indirect ruled areas seem evenly balanced on these observable characteristics.

3 Data

This section describes the datasets that are utilized in the paper. We describe the sources of the datasets and our sample.

3.1 Project Announcement Data

We obtain data on project announcements from the CapEx database maintained by the Centre for Monitoring Indian Economy (CMIE). CMIE, a leading business information company in India, was established in 1976 and has been used extensively for India-based academic research. More recently this data has been employed in Alok and Ayyagari (2020) to study the association between the electoral cycle and the corporate investment cycle. The CapEx database serves as the source of annual report – Private Investment Growth and Prospects published by the Indian central bank – the Reserve Bank of India. All projects announced by private and public firms and government entities that cost more than ₹ 10 mn (≈ \$0.2 mn) are recorded in the database. CMIE collects project announcement information from multiple sources including the annual reports of firms, media reports, and government agencies. Although CapEx has sporadic coverage of historical project announcements, the data is systematically available from 1995. The dataset provides information on the firm announcing the project, location (district, and state) of the project, size of the project, and time of project announcement. This dataset also allows tracking of the project outcomes.

The data employed in this paper covers project announcements across all states for the universe of all firms in India spanning between fiscal year 1996 and 2018.[4] However, we include 17 states in our sample because the empirical strategy in the paper exploits a within state estimation. Hence, states with at least one direct and indirect ruled district are included in the sample.[5] Figure 2 plots the geographic distribution of the total amount (in ₹ billion) and number of projects announced by all firms. Project announcements, in both number and amount, exhibit a great degree of geographic

[4]A fiscal year in India starts on 1st of April and ends on 31st of March in the next calendar year. As an example, fiscal year 2018 starts on 04/01/2017 (mm/dd/yyyy) and ends on 03/31/2018. Fiscal year is a reasonable unit of time as private and government enterprises manage their annual budgets within a fiscal cycle.

[5]The states included in the sample are Andhra Pradesh, Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, Tamil Nadu, Telangana, Uttar Pradesh, Uttaranchal, and West Bengal. Telangana and Andhra Pradesh were one state until 2014.

dispersion within 427 districts in the sample. Appendix C.1 presents an overview of the sample in terms of its distribution over time, amount, and project type.

3.2 Historical Geography of British Rule

The historical data on the boundaries of British rule, precolonial states, dates of annexation, mode of annexation by the British and details of the death of kings in precolonial states comes from Iyer (2010). We verify the historical data through primary data sources including Baden-Powell (1892), Chakrabarti (1896), Hunter and Bartholomew (1908), and Lee-Warner (1910). Iyer (2010) provides a crosswalk between precolonial boundaries and districts of India as per the 1961 Census. We manually match the 1961 Indian districts to the districts of India in 2014. The detailed list of 2014 Indian districts is obtained from the Environmental Systems Research Institute (ESRI), Inc. The 1961 districts are matched to the 2014 districts using the mapping discussed in Kumar and Somanathan (2009), and current district websites. District websites often include a webpage that details the history of a district, and accounts for administrative splits and name changes. If the district website does not have a history page, we refer to the Wikipedia page and verify information about splits or name changes through government notices. If a 2014 district matches with more than one 1961 district, we restrict the match to the 2014 district that reports the greatest geographic area. The 2014 district names are used to identify the districts where projects are announced in the CapEx database. Lastly, we collect data on precolonial cotton production, religion of the ruler, and 1911 census demographics at the district-level from Lee (2017).

Figure 1 shows the colonial boundaries of the areas under direct and indirect rule with the 2014 administrative state boundaries. *Direct ruled areas* consist of regions under the direct administrative rule of the British until the Indian independence in 1947. *Indirect ruled areas* are the native states under the administrative control of Indian princes, with indirect control by the British. Indirect ruled areas were present in all parts of British India, with a relatively higher concentration in the western and the central regions. Table A.1 shows the distribution of British acquisition of direct ruled districts over time and by reason.[6] The majority of districts were acquired during the ring fence period primarily by active conquest. 38% of districts that were acquired during subordinate isolation were through the Doctrine of Lapse, wherein a territory came under direct

[6] *District* refers to the modern day districts as of 2014, hereafter.

British rule if the ruler of the territory died without a biological natural heir. Districts that were acquired under misrule, comprise 7% of the sample of districts under direct rule.

Additionally, we obtain data on bilateral trade from Donaldson (2018). This data reports the disaggregated physical quantities of commodities shipped between two trade blocks. A trade block spans three to five districts. Four of the trade blocks are major ports of colonial India: Bombay, Calcutta, Karachi, and Madras. The trade flow data presents the final shipments between two regions. Donaldson (2018) collects trade data from various historical annual, provincial publications. We use this data to study how the total volume of cotton trade changed through the turn of the 19th century.

4 Results

This section describes the empirical methodology and key results. Our results show that investment is concentrated within states and this geographic concentration of investment can be explained by the proportion of districts under direct British rule.

4.1 Aggregate Analysis

We begin our empirical analysis by examining the concentration of corporate investment. We choose the state as the unit of analysis as the law is fixed at the state level.

4.1.1 Geographic Concentration of Investment

We begin our empirical analysis by examining the geographic concentration of corporate investment. The HHI measures the geographic concentration of investment. Using project level announcement data, we identify the location and size of each project. We then compute the state-level investment Herfindahl-Hirschman Index (HHI) as a measure of within state investment concentration. It is computed by summing the squared share of investment in a district, relative to the state in a given year. We use state as the unit of analysis, as the legal framework under which investment occurs is fixed within a state. Figure 3 plots the time-series of aggregate investment HHI between 1996 and 2018. The weighted HHI, weighted by the share of investment in the state relative to the country, shows a stable trend throughout the period with an average (median) value of 0.266 (0.255). The level of geographic concentration increases in the HHI.

While the HHI measures the geographic concentration of investment, it does not imply concentration on its own. Therefore, we compare the observed HHIs to three benchmarks: number HHI (# HHI), aggregate area HHI (Area HHI), and population HHI (Pop HHI). First, number HHI is computed as the inverse of number of districts in the state - $\frac{1}{N}$, where N denotes the number of districts in the state. This measure assumes equal investment across all districts. Second, state-level area HHI is computed as the squared sum of the share of geographic area of districts within a state. The aggregate area HHI is computed by weighing each state-level area HHI with the area of the state relative to the country. This measure assumes that the investment in each district, relative to the state, is proportional to the area of the district, relative to the state. Third, state-level population HHI is computed as the squared sum of the share of 2001 district-level population within a state. The aggregate population HHI is computed by weighing each state-level population HHI with the population of the state relative to the country. This measure assumes that the investment in each district, relative to the state, is proportional to the population of the district, relative to the state. These benchmarks act as a baseline for the geographic concentration of investment in a frictionless spatial equilibrium. The geographic concentration of investment implied by the observed HHI is 20 percentage-point higher than either measure of the benchmark. Hence, the comparison of the observed HHI with either of the three benchmarks indicate that investment is spatially concentrated within states.

4.1.2 What Explains the Geographic Concentration in Investment?

This section documents the relationship between the geographic concentration of investment and direct British rule. Figure 4 presents a scatterplot of state-level geographic concentration of investment and the percent of districts under direct British rule in that state. State-level geographic concentration of investment is measured using the average state-level investment HHI between 1996 and 2018. The plot provides prima-facie evidence that the geographic concentration of investment increases with the proportion of districts under direct British rule.

Table 2 extends the analysis presented in Figure 4 in a regression framework after controlling for other state-level covariates. Columns (1) to (6) present the estimate of the regression of state-level average HHI on the percent of districts under direct British rule, while sequentially adding state-level covariates. Across all columns the estimate of interest, the coefficient associated with % Direct Rule, is negative and statistically significant at the 10% level. While the statistical significance of the estimate is low, given the number of observations, the addition of other state-level covariates does not change the estimate or the precision of the estimate, despite the model R^2 increasing from 13% to 60%. This indicates that despite the limited statistical power of the regression, the relationship between state-level geographic concentration of investment and the percent of districts under direct British rule is robust. Economically, the estimate indicates that a one standard deviation increase in the percent of districts under direct British rule, corresponding to a 28 percent-point increase, is associated with a 0.12-0.15 standard deviation increase in investment concentration, relative to the mean. This corresponds to a 8.5%-10.7% increase in investment concentration. Moreover, our key explanatory variable, percent of districts under direct British rule, can explain 13% of total variation in investment concentration. Overall, our results indicate that the percent of districts under direct British rule can explain investment concentration, and this relationship is robust and economically meaningful.

4.2 Baseline Analysis

Thus far, we have established a correlation between geographic concentration in corporate investment and percent of direct ruled districts. In this section, we provide a systematic analysis of this relationship using micro-level data from project announcements, showing that smaller projects are announced in districts that were historically under direct British rule. The key advantage of using granular data is to rule out potential biases that may create a wedge between the true and the observed relationship between investment and direct British rule.

4.2.1 Empirical Strategy

The objective of this section is to systematically analyze the relationship between project size and the historical origin of a district. We run a regression of the natural logarithm of the project size on a binary variable that takes a value of 1 if the district where the project is announced was historically under direct British rule, and 0 otherwise. Equation 1 presents the baseline specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct rule}_j + \theta_{i,y} + \theta_{s(j \in s),y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t} \tag{1}$$

$ln(Y_{i,j,t})$ denotes the natural logarithm of project size announced by firm i in district j, where district j is in state s, at time t in fiscal year y. Direct rule$_j$ is a binary variable taking a value of 1 if the district where the project is announced was historically under direct British rule. $\theta_{i,y}$ denotes firm $\times$ year fixed effects controlling for all time-varying observed and unobserved heterogeneity at the firm level. $\theta_{s(j\in s),y}$ denote state $\times$ year fixed effects controlling for all time-varying observed and unobserved heterogeneity at the state level. Specifically, $\theta_{s(j\in s),y}$ controls for the law at the state level. Additionally, θ_t denotes time fixed effects accounting for aggregate shocks and temporal trends in corporate investment. The primary source of heterogeneity in equation 1 relates to precolonial boundaries. Hence, we estimate the standard errors by two-way clustering at the precolonial boundaries and time level.[7] Additionally, we report Conley (1999) standard errors to ensure that our inference is not contaminated by spatial noise. We also control for the latitude and the longitude of the district to capture spatial trends (Kelly, 2019), spatial correlation in business cycles coming from the relation between distance to equator and economic performance (Acemoglu, Johnson and Robinson, 2001), and social infrastructure (Hall and Jones, 1999).

We interpret β in equation 1 as a within-state estimator, estimating the difference in investment size of projects between direct and indirect ruled districts within a state, while controlling for firm-specific time-varying policy. Alternatively, β may be interpreted as a within-firm estimator, controlling for time-varying changes in the legal framework within a state. Under the second interpretation, the estimate gives the log difference in the project size announced by a firm in a direct ruled district relative to an indirect ruled district.

4.2.2 Baseline Results

Table 3 reports the regression results from the estimation of equation 1. The coefficient of interest is the estimate β associated with Direct Rule$_j$. Columns (1) to (5) sequentially add fixed effects to estimate equation 1, as reported in column (5). The estimate of β is negative and statistically significant across all columns. The magnitude of the estimate in column (5) indicates that projects announced in direct ruled districts are 8.8% smaller in size than projects announced in indirect ruled districts by the same firm, while controlling for the legal framework. The estimate indicates that for every project worth ₹ 100 million in an indirect ruled district, the size of the project in a direct ruled

[7] The total number of clusters with precolonial boundaries are 95, and 92 with time in the sample.

district is ₹ 91.19 million. This amounts to an average within-state difference of ₹ 8.81 million per project across districts.[8] Given the median project size of ₹ 1000 million, this corresponds to a within-state difference of ₹ 88.1 million (≈ $1.2 million) per project on annual basis, corresponding to a difference of $120 million per 100 projects. We further discuss and elaborate on the economic significance of the estimates in section 4.2.3.

Overall, the results indicate the estimate is negative and statistically significant.[9] The negative estimate of β indicates that districts that were historically under direct British rule exhibit lower investment today relative to districts under indirect rule.

4.2.3 Extensive and Intensive Margin

This section presents the results on the relation between investment and direct British rule, using district as the unit of analysis. The motivation to conduct the analysis at the aggregate level is threefold. First, aggregating data at the district-time level allows us to create a balanced district-time panel. This facilitates the quantification of the spatial differences in corporate investment via the extensive and the intensive margins. Second, this allows benchmarking the magnitude of the estimate, making it easier to interpret the estimates. Third, the key level of heterogeneity we exploit is at the district level whereas the key dependent variable is at the firm × district level. Moulton (1990) argues that a regression estimating the effect of an aggregate variable on a micro unit can lead to standard errors that are downward biased resulting in type I error.

Table 4 reports the results on the relation between corporate investment and direct British rule at the district-level. The dependent variable in column (1) takes a value of 1 if a project is announced in that district-year and 0 otherwise. Column (2) uses total investment in district j as the dependent variable. Column (3) uses total investment in district j, conditional on positive investment as the dependent variable. Column (4) uses the total number of projects announced in district j. Column (5) uses the total number of projects announced, conditional on announcement in district j. Column (6) uses the share of investment in district j relative to investment across all districts in the state. Column (7) uses the share of the number of projects relative to the total number of projects announced in the state as the dependent variable. We control for state × time fixed effects

[8]The effect is calculated using $\mathbb{E}[ln(Y_{i,1,t})|\text{Direct Rule}_j = 1, j \in s] - \mathbb{E}[ln(Y_{i,0,t})|\text{Direct Rule}_j = 0, j \in s] = \beta$

[9]We also validate the precision of the estimates through the Conley (1999) standard errors, adjusting for spatial dependence within 100 km, reported in square brackets.

in all columns to account for observed and unobserved state-level time-varying heterogeneity. We interpret the estimate of direct rule (=1), as a within-state estimator.

Across all specifications, the coefficient of direct British rule is negative and statistically significant in both the extensive and the intensive margins of corporate investment. Using a linear probability model, in column (1), we find that projects are 25% less likely to be announced in direct ruled districts relative to indirect ruled districts within a state. Column (2) shows that total investment is ₹16 billion lower in direct ruled districts relative to indirect ruled districts within a state. On the intensive margin, we find that conditional on announcement, project size is lower by ₹28 billion in direct ruled districts relative to indirect ruled districts within a state, as shown in column (3).[10] In addition, on average, for five projects announced in indirect ruled districts, only one project is announced in direct ruled districts within the state, shown in column (4). Furthermore, conditional on announcement, the ratio of projects announced in indirect to direct ruled districts is 7:1, shown in column (5). Lastly, we find the share of investment and the share of number of projects are 7% lower in direct ruled districts compared to indirect ruled districts within a state, shown in columns (6) and (7). This estimate is economically large; on average, a district accounts for $\approx$ 9% of the states' total investment and project count. Overall, the results indicate that the long-run effect of direct British rule is negative, statistically significant, and economically large. Moreover, the effect shows up in both the extensive and the intensive margins of corporate investment.

4.2.4 Robustness – Omitted Variable Bias

While the estimate of β, reported in Table 3, is negative and statistically significant across all columns, the magnitude of the estimate decreases in comparing across columns, suggesting omitted variable bias may be a concern (Altonji, Elder and Taber, 2005; Oster, 2019). From column (1) to column (5), the magnitude of the point estimate of β decreases from -0.17 to -0.09 with a simultaneous increase in the model R^2 from 3% in column (1) to 72% in column (5). Instability

[10]The results are robust to using alternative measures of the dependent variables. Appendix Table D.1 presents the results. In columns 1 and 3 we use one plus the natural logarithm of district-level investment, and one plus the natural logarithm of the number of projects as the dependent variables, respectively. In columns 2 and 4, we use the natural logarithm of district-level investment, and the natural logarithm of the number of projects as the dependent variable, respectively. Further, in columns 5 and 6 we demonstrate robustness using a Poisson pseudo-maximum likelihood model as advised by Chen and Roth (2023), with the dependent variable in level form. We prefer the reported regressions over using other transformations or estimators for ease of the interpretation of the estimate.

in the estimate may suggest the presence of omitted variable bias. We address this concern by conducting an Oster (2019) test. The Oster (2019) identified set based on the change in β and the model R^2 between column (1) and column (5) is $(-0.0893, -0.0549)$, which safely excludes 0. Furthermore, we do a step-wise Oster (2019) test. The identified sets on moving from column (1) to column (2), column (2) to column (3), and column (3) to column (4) are (-0.113,-0.048), (-0.133,-0.115), and (-0.086, -0.036), respectively. All of the sets safely exclude 0. These results suggest that omitted variables are unlikely to drive the relationship between project size and historical origins of a district, under the assumption Oster (2019).

4.2.5 Robustness – Controlling for Other Geographic & Historical Covariates

This section examines the robustness of the baseline estimates to other time-invariant district-level covariates. The British were likely to take over strategic territories. This includes territories such as those along the coast, those with larger geographic areas, and those with favorable climatic conditions. Moreover, they could selectively choose territories based on the demographic characteristics of the rulers, elites, and populace in those regions. To account for this, we include geographic and historical covariates, which may be correlated with the likelihood of a district being under direct rule.

Appendix Table D.2 reports the estimation results of the baseline equation 1 after controlling for geographic covariates. Appendix Table D.3 presents the baseline results with the inclusion of historical covariates such as the ruling empire, religion of the precolonial ruler, and historical population demographics such as the proportion of Muslims, Sikhs, lower caste individuals, and elites. Our estimate of interest is the coefficient of Direct Rule$_j$, which remains negative, statistically significant and stable in magnitude. Overall, the results indicate that the estimate of β is unlikely to be driven by other geographic and historical characteristics that could potentially be correlated with the direct rule variable.

4.2.6 Robustness – Placebo Test

We conduct a placebo test wherein we randomize whether a district was under direct British rule or not. This test validates that the results are neither spurious nor driven by spatial autocorrelation of investment or other spatial noise (Kelly, 2019). We randomly assign a district to be under direct British rule or indirect British rule, irrespective of the colonial history of the district.[11] We estimate the baseline specification, equation 1, using the random assignment of districts into direct and indirect ruled districts as the key explanatory variable. We refer to this as placebo direct rule. We estimate the coefficient associated with the placebo direct rule and repeat this exercise 10,000 times. To invalidate the baseline results, the null hypothesis that the point estimate associated with placebo direct rule is zero must be rejected.

Figure 5 presents a visual assessment of the kernel density of β, coefficient associated with placebo direct rule, estimated using 10,000 simulations. The distribution of β is centered around 0, varying from -0.106 to 0.087 with a standard deviation of 0.026. We fail to reject the null hypothesis – the average point estimate from the placebo analysis is equal to zero. The dashed, red line denotes the location of the coefficient of the interaction term from column 5 of Table 3. 0.1% of estimates, among the 10,000 simulated placebo β, lie to the left of the dashed red line. The results of the placebo test corroborate the argument that the baseline results are neither spurious nor driven by spatial noise or spatial autocorrelation of investment or other spatial noise (Kelly, 2019).

5 Evidence from Bordering Districts

Our approach, so far, of comparing the differences in investment levels across direct and indirect ruled district implicitly assumes that one district in India is as good a control as any other district within a state. Such a comparison can bias our inference in the presence of systematic differences between direct and indirect ruled districts. This section addresses this concern by focusing on contiguous districts that are separated only by administrative borders, specifically in cases where one district was historically under direct British rule and another under indirect rule within a state. This sample of contiguous districts is shown in Figure 6a. The identifying assumption of using this subsample is that these adjacent districts are expected to have followed similar paths had India not been colonized. Moreover, whether a district within a contiguous direct-indirect ruled pair was

[11] Each district has a probability of 0.6370 to be assigned as directly ruled district. This probability level is chosen based on the observed empirical probability of being a direct ruled district in the data. The random assignment of districts ignores the spatial autocorrelation in the likelihood of a district being under direct or indirect British rule but retains the spatial autocorrelation in investment. Ignoring the spatial autocorrelation in the likelihood of direct rule is of little concern as the Moran I for direct rule is small indicating that the direct rule variable is relatively randomly distributed in the space. Section E discusses the calculation and the magnitude of spatial autocorrelation of direct British rule.

under direct British rule or not is likely to be a matter of chance. Hence, the indirect ruled districts provide a valid counterfactual to the adjacent direct ruled districts. More importantly, focusing on this subsample of data allows us to implicitly control for the costs of moving goods, people, and ideas and geography – the four important determinants of investment choice (Marshall, 1920; Ellison and Glaeser, 1997; Ellison, Glaeser and Kerr, 2010) and investment opportunities, as these parameters are likely to be similar across contiguous districts.

5.1 Empirical Strategy

We combine the sample of contiguous direct and indirect ruled districts with a simple OLS regression framework to compare the difference in the size of investment projects across the two districts. We estimate the following regression specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,p(j\in p),y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t} \qquad (2)$$

where, $ln(Y_{i,j,t})$ denotes the natural logarithm of the size of the project announced by firm i in district j. Direct Rule_j is a binary variable taking a value of 1 if the district where the project is announced was historically under direct British rule, and 0 otherwise. As in the baseline specification 1, we also control for the latitude and longitude of the district. $\theta_{i,p(j\in p),y}$ denotes district-pair $\times$ firm $\times$ year fixed effects. District-pair refers to the adjoining pair of a indirect ruled district and a direct ruled district, within a state. This definition of district-pair can result in the presence of a single district in multiple pairs inducing a mechanical correlation across district-pairs. We account for multiple sources of correlations in the error term by multi-way clustering at the precolonial state (64 clusters), district-pair (174 clusters) and time (92 clusters) levels. The estimate of β is identified through within district-pair variation of the same firm, i.e., we effectively compare the size of investment projects announced by a firm in a direct ruled district with the size of investment projects announced by the same firm in bordering indirect ruled districts. This estimate pools all local comparisons and allows for spatial autocorrelation, addressing the dual problems of omitted variable bias and bias in the estimated standard errors.

5.2 Results from Bordering Districts

Table 5 reports the regression results from the estimation of equation 2 using the subsample of contiguous direct and indirect ruled districts. The coefficient of interest is the estimate of β associated with Direct Rule$_j$. Columns (1) to (3) sequentially add fixed effects to estimate equation 2, as reported in column (3). The estimate of β across all columns is negative and statistically significant. The point estimate in column (3) is marginally greater in magnitude relative to the baseline estimate reported in column (5) of Table 3. This indicates that any systematic differences across districts that are likely to affect the estimate of β in the traditional approach employed in equation 1 are likely to understate the true effect. The magnitude of the estimate in column (3) indicates that projects announced in direct ruled districts are 10.8% smaller in size than projects announced in indirect ruled districts by the same firm within a district-pair.

An important assumption of this local identification approach is that contiguous districts are likely to be similar to each other. While a simple balance test on some observed characteristics can provide reasonable confidence (see, Table 1), it is likely to be insufficient as it cannot rule out differences across the entire set of observed and, importantly, unobserved characteristics. We verify this assumption using a different approach that is better adept at accounting for observed and unobserved characteristics. We compare a new set of indirect ruled districts that are directly adjacent to the sample of indirect ruled districts, shown in Figure 6a. We refer to this new set of indirect ruled districts as "hinterland" indirect districts as these districts are separated from the sample of direct ruled districts in Figure 6a by a strip of indirect ruled districts. The rationale of this test is that if our original test comparing contiguous direct and indirect ruled districts is plagued by systematic differences across contiguous districts, then a comparison of the set of contiguous indirect ruled districts should also generate an effect of a similar magnitude. Figure 6b presents the sample. The indirect ruled districts, referred to as "border" indirect districts, used in Figure 6a are marked in blue and the contiguous indirect ruled districts, the hinterland indirect districts, are marked in gray. Table 6 reports the estimation results of comparing the size of investment projects within the sample of contiguous indirect ruled districts. The estimate of $Hinterland(=1)$ is economically small and statistically insignificant. This indicates that the results reported in Table 5 are unlikely to be driven by systematic differences across contiguous districts.

5.2.1 Project Abandonment

The ideal empirical design to test for lower investment in direct ruled districts should indicate whether a firm forgoes positive net present value (NPV) projects in direct ruled areas. However, the data for such an analysis is unavailable. Instead, we focus on the status of a project after announcement. Assuming that announced projects are positive NPV projects, the abandonment of these projects can reflect that a firm forgoes a positive NPV opportunities. This section examines the relationship between direct rule and the status of announced projects. Specifically, we compare the likelihood that a project is abandoned in direct ruled districts compared to indirect ruled districts.

We examine the differences in project post-announcement status by comparing the outcomes of a cohort of projects announced in a year. Specifically, we analyze the abandonment status within one, two, three, four, and five years of an announcement. This cohort analysis allows us to circumvent econometric issues related to the right-censoring bias. Moreover, this analysis allows us to address another concern. Specifically, the concern is that a direct comparison of the ex-post abandonment status of announced projects in direct and indirect ruled districts may not capture firms' tendency to forgo ex-ante positive NPV opportunities. Instead, the ex-post project abandonment could potentially reflect subsequent geographic or firm-specific events. We address this concern through our local identification approach, by including firm $\times$ district-pair $\times$ year fixed effects. The identifying assumption is that a firm is likely to have identical exposure to shocks within a pair of adjacent districts.

Table 7 reports the results from the linear probability model, combining the cohort analysis with the local identification approach. The table examines the probability that a given firm abandons a project in a direct ruled district relative to a contiguous indirect ruled district within the five years following the announcement. Columns (1)-(5) compare the likelihood of project abandonment within one, two, three, four, and five years of announcement, respectively. Our results show that the same firm is 2.5 percentage-points more likely to abandon its projects in a direct ruled district relative to a contiguous indirect ruled district. This effect is economically significant, given that the unconditional probability of project abandonment is 13%. These results indicate that firms may be more likely to forgo economically sound investment opportunities in direct ruled districts.

6 Instrumental Variable Regression

This section reports and discusses the results from an instrumental variable (IV) strategy to further address the selection issue. Thus far, we have compared corporate investment between districts that are under direct and indirect rule within a state. A concern with such a comparison is the problem of selection. While the low value of the Moran I statistic implies spatial randomness under a parametric setup, it does not completely rule out selection. We address the issue of selection by exploiting a unique feature of British annexation policy in India – the Doctrine of Lapse.

We use the British policy of the Doctrine of Lapse as an instrument for direct British rule. Under the Doctrine of Lapse, the governor-general of British India annexed the Indian princely states where the ruler died without a natural heir. This policy was in effect between 1848 and 1856. The Doctrine of Lapse allows us to exploit the death of a ruler without a natural heir, during 1848 to 1856, as an instrument for a territory coming under the direct British rule.

The instrument relies on two crucial assumptions: relevance and exclusion. The relevance restriction is that the death of a ruler without a natural heir can explain the likelihood of a territory coming under direct British rule. The relevance condition is a mechanical byproduct of the policy, which we verify in the first stage.[12] The exclusion restriction is that the death of a ruler without a natural heir affects current corporate investment only through the territory being under direct British rule. The identifying assumption required for the exclusion restriction to hold is that the death of the ruler without a natural heir during the period of 1848 to 1856 is likely a matter of chance and unlikely to have a direct impact on corporate investment in the present. We address issues related to the exclusion restriction in section 6.1.

Table 8 reports the results from the 2SLS regression and compares the estimate with the IV and the OLS estimate. The sample is restricted to all indirect ruled districts and districts under direct rule that were annexed after 1847. We restrict the direct ruled districts sample, because we are unable to construct the instrument for periods before 1848 as the policy was not applicable during other periods. Column (1) reports the OLS estimate for the sample under consideration. The estimate of β is negative and statistically significant. Column (2) presents the IV specification

[12] Lord Dalhousie was the governor-general of India between 1848-56. During his reign, the British gained direct control of seven native states wherein four were occupied by exercising the Doctrine of Lapse and three were annexed due to other reasons. In total, rulers of eight native states died during this period without a natural heir. This accounts for a total of 48 current districts being brought under direct British rule with 54.17% of districts annexed under the policy of Doctrine of Lapse.

regressing the natural logarithm of project size on the instrument. The IV estimate is negative and statistically significant. Columns (3) and (4) report the second stage and the first stage of the 2SLS estimation strategy, respectively. The instrument is positively related to the likelihood that a district fell under direct British rule. A district is 66% more likely to be under direct British rule if its ruler died without a natural heir between 1848 and 1856. The f-statistic of the first stage is 28.94 implying that the instrument is not a weak instrument and unlikely to overestimate the effect in the second stage. The second stage estimate of direct rule is negative and statistically significant at the 5% level. The magnitude of the 2SLS estimate in column (3) is comparable to the size of the OLS estimate in column (1), indicating that selection is likely to be of little concern. The closeness in the magnitudes of the OLS and the 2SLS estimates is consistent with very small values of the Moran I statistic, indicating that direct and indirect ruled districts may be randomly distributed in space. Overall, this analysis indicates that the results discussed so far are unlikely to be driven by selection bias.

6.1 Discussion on the Exclusion Restriction

This section discusses the validity of the exclusion restriction of the instrument. The exclusion restriction states that the death of a ruler without a natural heir affects current corporate investment only through the territory being under direct British rule. The identifying assumption required for the exclusion restriction to hold is that the death of the ruler without a natural heir between 1848 and 1856 is likely a matter of chance, and does not directly impact present day investment. In this section, we examine various challenges to the exclusion restriction, such as the strategic acquisition of territory by the British, voluntary withdrawal from the policy by incumbent rulers, and confounding variables that result in a ruler's demise without a natural heir. These factors could potentially impact the long-term consequences of a region through channels other than direct rule.

First, we argue that the Doctrine of Lapse was not strategically implemented to acquire certain territories, nor could rulers voluntarily opt out of this policy. Rahim (1963) notes that the Doctrine of Lapse did not allow the British to take over their most desired states of Oudh and Hyderabad. The British could resort to other modes of annexation to acquire their most desired states. Indeed, the British acquired Oudh and Hyderabad through other means. Another threat to identification is the possibility that the British were responsible for the deaths of the rulers whose kingdoms they wanted to acquire. Iyer (2010) reviews historical records and argues that neither case is likely. Moreover, the policy of lapse was completely unexpected, ruling out any self-selection into or out of the policy (Iyer, 2010). Additionally, all states were administered identically under direct British rule, irrespective of the mode of annexation.

Second, we rule out alternative hypotheses that may be driving our findings. For instance, the death of a ruler without a natural heir may be attributable to several other factors, e.g., environment, genetics, etc., which may affect postcolonial corporate investment through other channels. Alternatively, the death of a ruler without a natural heir may usher in a period of political turmoil, resulting in adverse long-run consequences captured by our estimation. We address these two claims via a falsification test. We directly regress the natural logarithm of project size on a binary variable that takes a value of 1 if the ruler died without a natural heir in the periods before 1847 and after 1856, in a sample of indirect ruled districts. The Doctrine of Lapse was not applicable in these periods and indirect districts were not subject to direct British rule. Hence, if the death of a ruler without a natural heir affects present day outcomes directly through channels other than the direct rule, we should find significant, comparable results in this sample. Column 5 of Table 8 reports the results. The coefficient of interest is close to zero and statistically insignificant. This indicates that the death of a ruler without a natural heir is unlikely to directly affect long-term outcomes.

Third, we show that the death of a ruler is unlikely to directly affect long-term outcomes. One concern with regard to our exclusion assumption is that the death of the ruler between 1848 and 1856, regardless of presence of natural heir, may reflect geography-specific conditions during the period that can affect long-term outcomes through channels other than direct rule. We test this proposition by directly regressing the natural logarithm of project size on a binary variable that takes a value of 1 if the ruler died *with* a natural heir in the periods after 1847 and before 1857, in a sample of indirect ruled districts. As before, these territories were not under direct rule, hence, we should find significant, comparable results in this sample if the death of a ruler during this period affects present day outcomes directly through other channels. Column 6 of Table 8 reports the results. The estimate of interest is both economically small and statistically insignificant. This indicates that the death of a ruler is unlikely to directly affect long-term outcomes.

7 Mechanism

This section discusses a mechanism through which direct British rule affects corporate investment in the present: destruction of existing economic organizations. We show this in the context of cotton production in India, a well-established industry that posed a significant threat to the British textile industry, before the onset of British colonial rule. We begin by discussing the motivation for capture of the cotton industry and documenting precolonial cotton production as a strong determinant of direct British rule. Further, we use the adoption of the Whitney cotton gin as a natural experiment to examine the impact of shifts in British incentives on their pursuit of cotton in India. We show that there are long-run effects associated with the destruction of the cotton industry, and explain the ways through which colonial policies and practices have persisted to the present.

7.1 Capture of the Cotton Industry

> *England began with driving the Indian cottons from the European market; it then introduced twist into Hindostan, and in the end inundated the very mother country of cotton with cottons.*

-Marx (1853)

India produced about 25% of the world's manufacturing output in 1750, of which, textiles constituted a significant share (Marks, 2019). Indian textiles dominated the world textile market in the 18$^{\text{th}}$ century, accounting for 25% of the global textile trade (Maddison et al., 1995). The Indian cotton textiles were the most important manufactured goods in the 18$^{\text{th}}$ century with India being home to the world's most important cotton textile industry (Parthasarathi, 2011; Robson, 1957). Cypher (2008) notes that the Indian textile production was marked by the presence of skilled laborers and large factory towns, which threatened the British textile industry – a leading sector of the British economy.

Politically, the British, through the EIC, consolidated economic power in India by annexing cotton producing territories (Schoen, 2009). The direct control of cotton producing areas allowed the British to control the supply of cotton, a valued commodity with a rising market share in the textile industry. The control of the cotton allowed the British to meet two objectives: (1) protect the interests of the domestic textile industry in Britain, and (2) increase Britain's share of global trade, by directly controlling India's largest industry. This is also reflected in the historical narrative that the objective of the East India Company was to to make India an agricultural colony of British capitalism by securing a monopoly on the supply of Indian goods and products, specifically cotton (Sahoo, 2015).

The negative effects of the destruction of India's textile industry is evidenced by its fall in global textile exports, from 25% in the 17th century to just 2% at the end of British rule in 1947 (Das, 1946). The destabilization of the well-flourishing textile industry resulted from the policies and practices of the British. We emphasize four ways through which the textile industry, specifically the cotton industry, was destroyed in India: (1) protectionism of the British textile industry through tariffs, (2) price fixing through monopsony power, (3) violence against textile producers, and (4) deprivation of new technological innovations in the local industries. We direct readers to appendix section B for a detailed discussion on the ways through which the Indian textile industry was destroyed, specifically the cotton industry. Overall the attitude of the British towards the Indian cotton textile production is best described by HH Wilson, in his continuation of James Mill's *The History of British India*.

> "... had not such prohibitory duties and decrees existed, the mills of Paisley and of Manchester would have been stopped in their outset, and could hardly have been again set in motion, even by the powers of steam. They were created by the sacrifice of the Indian manufactures. Had India been independent, she would have retaliated; would have imposed preventive duties upon British goods, and would thus have preserved her own productive industry from annihilation. This act of self-defence was not permitted her; she was at the mercy of the stranger. British goods were forced upon her without paying any duty; and the foreign manufacturer employed the arm of political injustice to keep down and ultimately strangle a competitor with whom he could not contend on equal terms" (Mill and Wilson, 1848).

We document empirical evidence, consistent with the historical narrative that the British took direct control of cotton-producing areas. Figure 7 shows that 80% of all districts under direct British rule were cotton producing districts. A cotton producing district was 40% more likely to be under direct British rule, relative to indirect rule. The likelihood that a cotton producing district is under direct British rule is economically large and statistically significant.

7.2 Destruction of Cotton Industry and Adoption of Whitney Cotton Gin

This section examines how British incentives to control the Indian cotton industry evolved in response to a significant shift in global cotton supply – the rise of the American colonies as the primary cotton supplier. This shift acted as a shock, reducing the incentives of the British to control and expropriate India's cotton sector, ultimately leading to reduced economic destruction within the domestic cotton industry for areas annexed later.

The emergence of the American colonies as a low-cost supplier of cotton by 1801 reduced the incentives of the British to continue their exploits of cotton in India.[13] This transformation began with Eli Whitney's patenting of the cotton gin in 1794 and its subsequent adoption in the US by 1800. The Whitney cotton gin effectively and efficiently removed the seeds from cotton plants, thereby making the process of extracting seeds from fibers for the short-staple type cotton very cheap. In contrast, the cotton farmed in India was of the long-staple variety and did not benefit from this invention. Logan (1965) notes a widespread adoption of the superior and cheap American short staple among Lancashire textile manufacturers. As a result, after the adoption of the Whitney cotton gin, Indian cotton fell out of favor with British cotton mills. After 1800, the British had less motive to expropriate the Indian cotton textile sector.

Combining the timing of the widespread adoption of Whitney cotton gin with data on Indian domestic exports of agricultural commodities 1884-1920, we document that the domestic cotton exports of provinces annexed before 1800 ("early") are lower than the domestic exports of provinces annexed after 1800 ("late"). The objective of this test is to provide evidence supporting the underlying mechanism of economic destruction. If other channels were driving the observed differences, we would not observe distinct patterns in trade flows of the cotton commodity between areas that were annexed early versus those annexed later. In the absence of changes in incentives for the capture of the cotton industry, we would expect uniformity in the impact across both groups. However, the clear differences in outcomes between early and late annexed regions suggest that economic destruction plays a pivotal role in shaping the observed results. This finding implies that the differential effects of economic disruptions are driven by changes in incentives for the capture of the cotton industry in India – rather than other potential factors – are responsible for the variations in outcomes across these regions. Therefore, the timing of annexation serves as a critical factor in understanding the extent and nature of economic destruction in these areas.

Table 9 presents this result. The table examines the relationship between the natural logarithm

[13] In 1801, the US's annual cotton production exceeded 22 million kg. By the early 1830s, the US had become the world's leading cotton producer (Oyangen, 2014).

of trade flows for each commodity between trade blocks on the interaction term of early annexation and cotton commodity. Early annexation takes a value of one for trade blocks that were under direct British rule before 1800, and zero, otherwise. Cotton commodity takes a value of one for trade flows of cotton and zero for all other commodities. Our preferred specification includes exporter province × importer province × commodity fixed effects, commodity × year fixed effects, and exporter block × importer block × year effects. Columns 1-6 sequentially add these fixed effects. The coefficient on the interaction term is negative and statistically significant across all columns. The negative coefficient indicates that the cotton exports of trade blocks that were annexed before 1800 were lower than the cotton exports of trade blocks annexed after 1800.

This result indicates that the negative economic effects of the British capture of the Indian cotton industry were concentrated among districts that were annexed before the widespread US adoption of the cotton gin. This finding is consistent with our hypothesis that after 1800, the British had weaker incentives to expropriate the Indian cotton textile sector with the rise of American cotton. We also note that this effect is not driven by a lower likelihood of cotton-producing areas being annexed under direct British rule after 1800. Figure 8 shows that among districts under direct rule, 80% of districts were cotton producing districts regardless of the period in which they were annexed.

Taken together, the two results indicate that there was no shift in the annexation policy, rather a shift in how the local cotton industry was treated after annexation.

7.3 Long-Run Effect

This section examines the long-run consequences of colonial destruction of the cotton industry. In particular, we study how direct British rule affects corporate investment in the present through the destruction of the cotton industry during the colonial era.

The analysis uses cotton cultivation during precolonial times as an instrument for direct British rule in a sample of bordering districts, which include firm × district-pair × year fixed effects.[14] Table 10 presents the long-run effect of the destruction of the Indian cotton industry on corporate investment in the present, through its impact on direct British rule. Column 1 reports the IV estimate. The IV estimate is negative and statistically significant, indicating that areas with

[14] We also report similar results from the full sample in Appendix Table D.4 for completeness.

greater cotton cultivation during the precolonial period have lower corporate investment in the present. Columns 2 and 3 report the second and first stage from the 2SLS estimation, respectively. The first stage estimate in column 3 is positive and statistically significant with an f-statistic of 47.4. This implies the relevance of the instrument and indicates that the second stage estimate is unlikely to be upward biased due to weak instrument concerns. Column 2 reports the results from the second stage. The second stage estimate is negative and statistically significant, indicating that a firm reduces its project size by 20% in direct ruled areas relative to indirect ruled areas, within a pair of adjacent districts.

The key identifying assumption of this analysis is that cotton cultivation during the precolonial period does not directly affect corporate investment in the present through other channels. The usage of adjacent district pairs and firm × district-pair × year fixed effects helps alleviate several threats to our identification strategy. Adjacent district pairs are likely to have similar geographic and climatic conditions, conducive to cotton cultivation. Hence, whether one district grew cotton and its bordering district did not is likely to be a matter of chance.

We further strengthen our claim for the exclusion assumption by conducting a falsification exercise. The falsification exercise focuses on a sample of adjacent indirect ruled district pairs, in which one district grew cotton during the precolonial period and the other did not. If cotton cultivation during the precolonial period affects present corporate investment either directly or indirectly through channels other than direct British rule, we should find differences across these adjacent district pairs that were never under direct British rule. Column 4 of Table 10 reports the results from this falsification test. The estimate of interest is close to zero and statistically insignificant. This indicates that precolonial cotton cultivation is unlikely to explain differences in present corporate investment directly or indirectly through channels other than direct British rule.

Next, we examine the impact on present day corporate investment among direct ruled areas that were annexed before 1800 relative to areas annexed after 1800. The objective of this test is to complement the analysis in section 7.2 in documenting that the reduced investment in direct ruled areas is primarily driven by cotton producing areas that were annexed before 1800, when the British may have had greater incentives to destroy the existing cotton industry. Table 11 presents this analysis, restricting the sample to direct ruled areas. Column 1 reports a negative coefficient on precolonial cotton production. Column 2 adds an indicator variable for early annexation to the specification in column 1. The estimate of precolonial cotton reduces in magnitude and becomes statistically insignificant, whereas the coefficient of early annexation is negative and statistically significant. Column 3 supplements the specification reported in column 2 by adding the interaction term of early annexation and precolonial cotton production. The estimates of precolonial cotton production and early annexation attenuates in magnitude and statistical significance, whereas the interaction term remains negative and statistically significant. This indicates that the present day lower investment in historically direct ruled areas is driven by cotton producing districts that were annexed before 1800.

7.3.1 Human Capital Accumulation as a Channel for Long-Run Effect

The long-run consequences of the destruction of existing economic organizations can be attributed to its effect on human capital accumulation. Chaudhuri (1978) argues that India's stronghold in the cotton industry, at its peak, can be explained by a large supply of skilled labor with specialized knowledge that was transferred over generations. The destruction of the cotton industry reduced the incentives of the cotton laborers to pass down their knowledge to future generations and disincentivized the subsequent generation from acquiring these skills. While one might expect such economic disruption to incentivize workers to develop new skills, the cotton textile workforce, predominantly from lower-caste communities, faced significant social and economic impediments that limited their ability to transition into new professions (Broadberry and Gupta, 2005).

This section provides empirical evidence in support of this hypothesis. We show that human capital acquisition is lower among lower-caste individuals in directly ruled districts that produced cotton during the pre-colonial period and were annexed before the invention of the Whitney gin. We utilize data from the National Demographic and Health Survey (DHS) for households in directly ruled districts, using education outcomes and self-reported interest in education as proxies for human capital acquisition. Specifically, we compare the education outcomes of lower-caste individuals in early-annexed cotton-producing districts with those in later-annexed districts. The analysis focuses on household heads, with the key coefficient derived from the triple interaction of low caste, early annexation, and precolonial cotton production.

Table 12 presents the results on educational attainment. The key dependent variable is a binary indicator equal to 1 if the household head has received higher education (post-secondary education) and 0 otherwise. Columns 1–6 estimate the coefficient of interest associated with the triple interaction term of low caste, early annexation, and precolonial cotton across various combinations of fixed effects. We successively add granular levels of fixed effects along with latitude and longitude controls, to account for regional and temporal variation across district-ruled districts. Particularly, column 6 includes district fixed effects to control for all factors that may have influenced certain districts' ability to grow cotton during the precolonial period and their likelihood of annexation at specific times. Additionally, we include native state × caste and present-day state × caste fixed effects in this column to control for caste-level characteristics at the state level.

The estimate for the triple interaction term is negative, remains stable and statistically significant across all columns. Our most conservative estimate suggests that lower-caste households in directly ruled areas that produced cotton during the precolonial period and were annexed before the invention of the Whitney gin are 11.3% less likely to attain higher education compared to their counterparts in directly ruled cotton-producing areas annexed after the invention of the Whitney gin in the present. This finding indicates that human capital formation remains persistently weaker in these regions relative to other areas with similar historical economic conditions but later annexation.

To further support these findings, we extend our analysis to additional education outcomes, including the likelihood of having no formal education, attaining primary or secondary education as the highest level, literacy, and total years of schooling, using the same sample. The results, presented in Appendix Table D.5, suggest significant disparities. Lower-caste households in early-annexed, cotton-producing districts are 69% more likely to have no formal education, 26% less likely to complete primary school, 32% less likely to complete secondary school, 69% less likely to be literate, and have an average of 5.1 fewer years of schooling compared to their counterparts in later-annexed cotton-producing districts in the present. These findings highlight the persistent effect of economic destruction on educational attainment.

Our hypothesis on human capital acquisition posits that lower-caste households lose the community knowledge accumulated over generations due to the large-scale destruction of key economic structures. Furthermore, because these households belong to marginalized castes, they face significant barriers to entering new professions, which may ultimately erode their interest in acquiring education. To test this conjecture, we leverage survey data, specifically examining the variable "main reason never went to school." The dependent variable is a binary variable that takes a value of 1 if the the primary reason a household member did not attend school was a perceived lack of necessity or interest in education, and 0 otherwise.

Table 13 presents the results. As in prior analyses, the coefficient of interest is the interaction term for low caste, early annexation, and precolonial cotton. This captures differences in the stated interest in acquiring education between low-caste households in early-annexed cotton-producing districts and those in later-annexed cotton-producing districts. The estimate of interest is positive, statistically significant, and robust across various fixed effects specifications in columns 1-6. Our most conservative estimate indicates that lower-caste households in directly ruled areas that produced cotton during the pre-colonial period but were annexed before the invention of the Whitney gin are 18.5% more likely to report a lack of interest in acquiring education compared to their counterparts in later-annexed districts.

Overall, these results shed light on the underlying mechanism behind the long-run effects of economic destruction. Specifically, we document that economic destruction can have long-term effects by diminishing the motivation to invest in human capital, particularly among marginalized groups that face structural barriers to mobility and adaptation.

7.3.2 Alternative Explanations: Public Goods, Trust & Conflict

This section discusses three alternative explanations that can potentially explain our results – differences in law and enforcement, provision of public goods, and trust. We argue that none of these forces can account for our results.

Law and Enforcement: An alternative mechanism is that there were pre-existing differences in the rule of law across direct and indirect ruled districts that persist to date and can explain the differences in present-day investment. There are three specific concerns, related to law and enforcement. We rule out all three.

The first potential concern may be that the British faced a high mortality rate in India, discouraging them from active investments in improving the quality of institutions (Acemoglu, Johnson and Robinson, 2001). Under this hypothesis, the differences in the rule of law across direct and indirect ruled districts could potentially explain differences in investment. However, this hypothesis is unlikely to explain our results as our estimation strategy compares direct and indirect ruled districts within a state. The law governing contract enforcement, property rights, and dispute resolution is fixed within a state regardless of the historical origins of the district.[15]

The second potential concern may be that although the law may be fixed within a state, the historical origins can affect the enforcement of the law across districts. This weak enforcement can negatively affect investment as documented by Ponticelli and Alencar (2016). This channel is also unlikely to explain our results as all corporate cases are adjudicated by the respective high courts or debt recovery tribunals of the state rather than the judicial machinery at the district level. Therefore, our within-state estimation strategy ensures that all firms are not only operating in an identical legal framework but are also exposed to identical enforcement of the legal framework.

A third potential concern may be that differences in enforcement at the district level can affect projects indirectly through other channels. We rule out this hypothesis by examining differences in court delays, the average time taken by local courts to deliver a verdict for criminal cases, across district types. Appendix Table F.1 shows that the differences in court delays across district types are insignificant for our sample. Hence, our results cannot be explained by differences in local enforcement across district types.

Provision of Public Goods: We examine the differences in the provision of public goods across direct and indirect ruled districts. Using data on public goods provision across Indian districts from 1961 until 1991, Iyer (2010) documents the differences across direct and indirect ruled districts. The differences in the provision of public goods diminish substantially over time, exhibiting convergence, consistent with policies of equalization followed by postcolonial governments (Iyer, 2010; Banerjee and Somanathan, 2007). Using the Census data for 1991, 2001 and 2011, we verify that the differences in public goods provision, measured by number of schools and electricity supply, are insignificant for our sample (see Appendix Table F.2). Hence, the differences in the provision of public goods are unlikely to drive our results.

[15] Moreover, it is worth noting that India is an outlier to the Acemoglu, Johnson and Robinson (2001) hypothesis. See Appendix Figure F.1. Figure F.1a presents a scatter plot of the average protection against expropriation and settler mortality, after standardizing the value for both the variables to zero for India for the Acemoglu, Johnson and Robinson (2001) sample. The plot shows that only five countries - Australia, Canada, New Zealand, Singapore, and United States have average protection against expropriation superior to India. Additionally, the three neighboring countries, Bangladesh, Sri Lanka (direct ruled) and Pakistan (partly direct ruled) have similar settler mortality to India but inferior protection against expropriation. The regression of economic prosperity, measured as the natural logarithm of GDP per capita (PPP) in 1995, on average protection against expropriation in Figure F.1b reinforces that India is an outlier. Hence, the theory of origins of formal institutions as posited in Acemoglu, Johnson and Robinson (2001) is unlikely to explain the institutions and the relationship between institutions and economic growth in India.

Trust: We have argued that direct ruled districts were under the direct administration of the British, hence, more prone to exploitation. This is likely to have fomented greater mistrust in the state during colonial rule in direct ruled districts compared to indirect ruled districts. This mistrust in the state can persist over the long-term (Nunn and Wantchekon, 2011). The lack of trust in the state and its machinery can impede the ability of the state from investing in basic infrastructure required for corporate investment. Therefore, an alternative hypothesis is that the differences in investment in the present reflect differences in trust in the state, attributable to the colonial origins of districts. Using the 2011 survey data from the Indian Human Development Survey (IHDS) on trust in state agents, such as politicians, police, panchayats, and banks, we do not find any evidence of systematic differences in trust in state actors across direct and indirect ruled districts (see, Appendix Figure F.2). Hence, we argue that differences in trust, due to colonial rule, are unlikely to drive our results. This observation of lack of differences in trust between direct and indirect ruled areas is consistent with Iyer (2010). Using data on state-level elections from the 1960s and from the 1980s, Iyer (2010) finds that there are no significant differences in voter turnout or the vote margin of the winning candidates, a proxy for the competitiveness of elections and trust in the incumbent state leader.

8 Conclusion

In this paper, we show that history can explain geographic concentration of investment over and above the traditional agglomerative forces, geographic differences and expectations. This paper uses within country variation in historical circumstances, combined with a local identification approach and instrumental variable strategy to explain the spatial differences in investment. We use spatial variation in direct and indirect British rule to identify differences in historical circumstances. Our aggregate analysis shows that the differences in historical origins can explain 13% of total geographic variation in investment. Our micro-level estimate suggests that investment is 8-10% lower in direct ruled areas, relative to indirect ruled areas. We further explore the channel through which history affects investment – economic organizations. We show that cotton-producing districts are more likely to be under direct British rule, and subject to adverse economic policies, resulting in the destruction of existing economic organizations with long-run detrimental effects.

Our work has three distinct contributions. First, we focus on a novel aspect that can explain the geographic concentration of investment: history. Second, our empirical strategy is better adept at solving the difficult identification problems associated with identifying the relationship between history and economic activity. Third, we explain how differences in colonial rule can produce long-run consequences, through the destruction of well-established economic organizations.

Our findings demonstrate that history can have enduring influence on the trajectory of economic development within a country. More broadly, this work informs discussions on the root causes of inequality, aiding our understanding of how the vestiges of history can create cleavages within a nation. Future study on how historical processes can perpetuate inequality may be a fruitful area of work, to further the discussion on economic disparities.

References

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

Alavi, Hamza. 1982. *Capitalism and colonial production.* Routledge.

Allen, Charles, and Sharada Dwivedi. 1984. *Lives of the Indian princes.* Century.

Alok, Shashwat, and Meghana Ayyagari. 2020. "Politics, state ownership, and corporate investments." *The Review of Financial Studies*, 33(7): 3031–3087.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113(1): 151–184.

Arthur, W Brian. 1986. *Industry location patterns and the importance of history.* Center for Economic Policy Research, Stanford University.

Ash, Elliot, Sam Asher, Aditi Bhowmick, Daniel Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi. 2021. "Measuring Gender and Religious Bias in the Indian Judiciary." Working Paper.

Ashton, Stephen Richard. 1977. "British Policy Towards the Indian States, 1905-1939." PhD diss. School of Oriental and African Studies (University of London).

Avineri, Shlomo. 1969. *Karl Marx on Colonialism and Modernization: his despatches and other writings on China, India, Mexico, the Middle East and North Africa.* Anchor Books.

Baden-Powell, Baden Henry. 1892. *The land-systems of British India.* Vol. 2, Clarendon Press.

Banerjee, Abhijit, and Lakshmi Iyer. 2005. "History, institutions, and economic performance: The legacy of colonial land tenure systems in India." *American Economic Review*, 95(4): 1190–1213.

Banerjee, Abhijit, and Rohini Somanathan. 2007. "The political economy of public goods: Some evidence from India." *Journal of development Economics*, 82(2): 287–314.

Berger, Daniel. 2009. "Taxes, Institutions and Local Governance: Evidence from a Natural Experiment in Colonial Nigeria."

Bobonis, Gustavo J, and Peter M Morrow. 2014. "Labor coercion and the accumulation of human capital." *Journal of Development Economics*, 108: 32–53.

Bolts, William. 1775. *Considerations on India Affairs: Particularly Respecting the Present State of Bengal and Its Dependencies...* Vol. 3, J. Almon.

Broadberry, Stephen N, and Bishnupriya Gupta. 2005. "Cotton textiles and the great divergence: Lancashire, India and shifting competitive advantage, 1600-1850."

Chakrabarti, Jadab Chandra. 1896. *The native states of India.* Luzac & Company.

Chaudhuri, Kirti N. 1978. *The Trading World of Asia and the English East India Company: 1660-1760.* Cambridge University Press.

Chen, Jiafeng, and Jonathan Roth. 2023. "Logs with zeros? Some problems and solutions." *The Quarterly Journal of Economics*, qjad054.

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

Cypher, James. 2008. *The process of economic development.* Routledge.

Das, Taraknath. 1946. "The Economic History of India: 1600-1800."

Dell, Melissa. 2010. "The persistent effects of Peru's mining mita." *Econometrica*, 78(6): 1863–1903.

Dell, Melissa, and Benjamin A Olken. 2020. "The development effects of the extractive colonial economy: The dutch cultivation system in java." *The Review of Economic Studies*, 87(1): 164–203.

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

Duranton, Gilles, and Henry G Overman. 2005. "Testing for localization using micro-geographic data." *The Review of Economic Studies*, 72(4): 1077–1106.

Ellison, Glenn, and Edward L Glaeser. 1997. "Geographic concentration in US manufacturing industries: a dartboard approach." *Journal of Political Economy*, 105(5): 889–927.

Ellison, Glenn, Edward L Glaeser, and William R Kerr. 2010. "What causes industry agglomeration? Evidence from coagglomeration patterns." *American Economic Review*, 100(3): 1195–1213.

Ethier, Wilfred J. 1982. "Decreasing costs in international trade and Frank Graham's argument for protection." *Econometrica: Journal of the Econometric Society*, 1243–1268.

Gandhi, Mohandas Karamchand. 1921. "Notes." *Young India*, 19: 487.

Gopal, Sarvepalli. 1965. "British Policy in India, 1858-1905." *Cambridge*.

Hall, Robert E, and Charles I Jones. 1999. "Why do some countries produce so much more output per worker than others?" *The Quarterly Journal of Economics*, 114(1): 83–116.

Huillery, Elise. 2009. "History matters: The long-term impact of colonial public investments in French West Africa." *American Economic Journal: Applied Economics*, 1(2): 176–215.

Hunter, William Wilson, and John George Bartholomew. 1908. *The Imperial Gazetteer of India.* Vol. 22, Clarendon Press.

Iyer, Lakshmi. 2010. "Direct versus indirect colonial rule in India: Long-term consequences." *The Review of Economics and Statistics*, 92(4): 693–713.

Kelly, Morgan. 2019. "The standard errors of persistence."

Krugman, Paul. 1987. "The narrow moving band, the Dutch disease, and the competitive consequences of Mrs. Thatcher: Notes on trade in the presence of dynamic scale economies." *Journal of development Economics*, 27(1-2): 41–55.

Krugman, Paul. 1991. "History versus expectations." *The Quarterly Journal of Economics*, 106(2): 651–667.

Kumar, Hemanshu, and Rohini Somanathan. 2009. "Mapping Indian districts across census years, 1971-2001." *Economic and Political Weekly*, 69–73.

Lee, Alexander. 2017. "Redistributive colonialism: The long term legacy of international conflict in India." *Politics & Society*, 45(2): 173–224.

Lee-Warner, William. 1910. *The Native States of India.* Macmillan and Company, limited.

Logan, Frenise A. 1965. "India's Loss of the British Cotton Market After 1865." *The Journal of Southern History*, 31(1): 40–50.

Lowes, Sara, and Eduardo Montero. 2020. "Concessions, violence, and indirect rule: Evidence from the congo free state." National Bureau of Economic Research.

Maddison, Angus, et al. 1995. *Monitoring the world economy, 1820-1992.* Development Centre of the Organisation for Economic Co-operation and . . . .

Marks, Robert B. 2019. *The origins of the modern world: A global and environmental narrative from the fifteenth to the twenty-first century.* Rowman & Littlefield.

Marshall, Alfred. 1920. "Principles of Economics." *London: MacMillan*, 621.

Marx, Karl. 1853. *The british rule in India.* Electric Book.

Méndez-Chacón, Esteban, and Diana Van Patten. 2019. "Multinationals, monopsony and local development: Evidence from the united fruit company." *Unpublished Manuscript*.

Mill, James, and Horace Hayman Wilson. 1848. "History of British India. Vol. I."

Moulton, Brent R. 1990. "An illustration of a pitfall in estimating the effects of aggregate variables on micro units." *The Review of Economics and Statistics*, 334–338.

Mukund, Kanakalatha. 1992. "Indian textile industry in 17th and 18th Centuries: Structure, organisation and responses." *Economic and political weekly*, 2057–2065.

Nakatomi, K. 2008. *History of Indian Textiles.* K. Hatanaka, ed., Textile Arts of India. San Francisco: Chronicle Books.

Nunn, Nathan. 2008. "The long-term effects of Africa's slave trades." *The Quarterly Journal of Economics*, 123(1): 139–176.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The slave trade and the origins of mistrust in Africa." *American Economic Review*, 101(7): 3221–52.

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

Oyangen, Knut. 2014. "The cotton economy of the Old South." *American Agricultural History Primer*.

Panagariya, Arvind. 1986. "Increasing returns, dynamic stability, and international trade." *Journal of International Economics*, 20(1-2): 43–63.

Parthasarathi, Prasannan. 2011. *Why Europe grew rich and Asia did not: Global economic divergence, 1600–1850.* Cambridge University Press.

Ponticelli, Jacopo, and Leonardo S Alencar. 2016. "Court enforcement, bank loans, and firm investment: evidence from a bankruptcy reform in Brazil." *The Quarterly Journal of Economics*, 131(3): 1365–1413.

Porta, Rafael La, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert W Vishny. 1998. "Law and finance." *Journal of Political Economy*, 106(6): 1113–1155.

Rahim, MA. 1963. "Lord Dalhousie's Administration of the Conquered and Annexed States (New Delhi: S. Chand)."

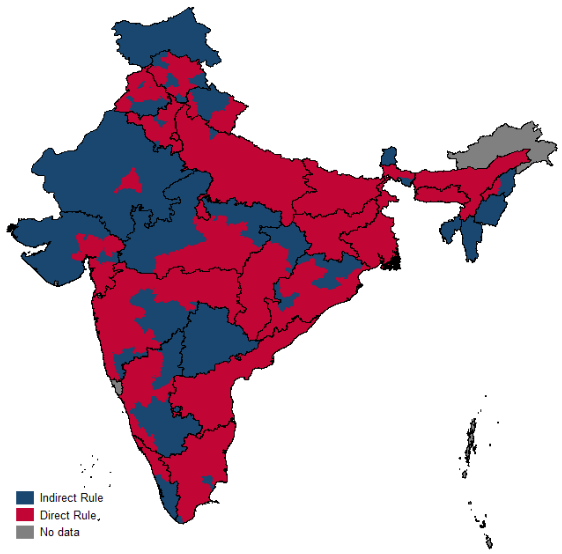
Robson, Robert. 1957. "cotton industry in Britain."

Sahoo, Rajib Lochan. 2015. "Indian Cotton Mills and the British Economic Policy, 1854-1894." *Proceedings of the Indian History Congress*, 76: 356–367.

Schoen, Brian. 2009. *The Fragile Fabric of Union: Cotton, Federal Politics, and the Global Origins of the Civil War.* JHU Press.

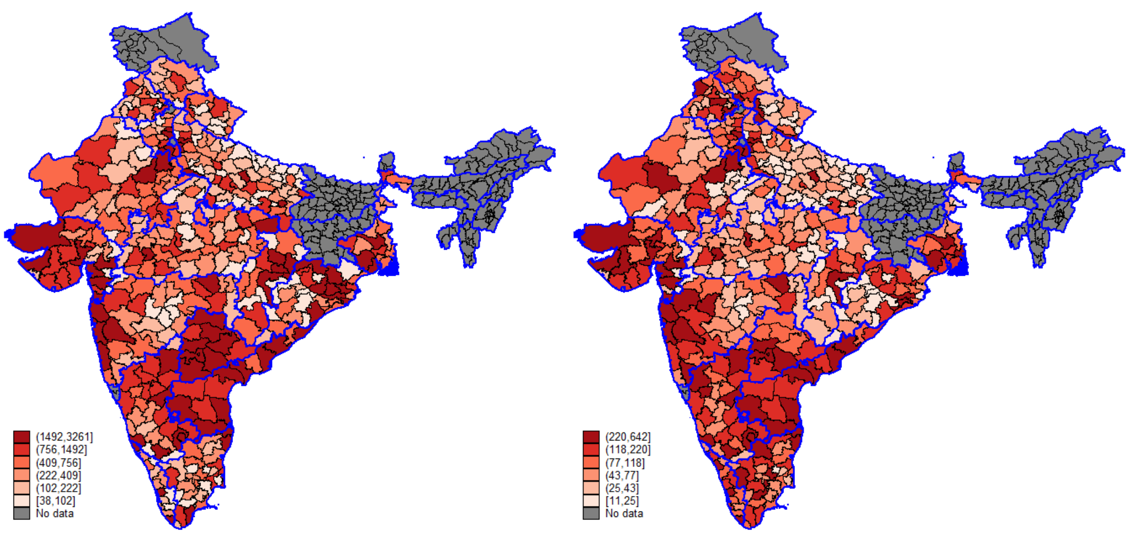
Zingales, Luigi. 2017. "Towards a political theory of the firm." *Journal of Economic Perspectives*, 31(3): 113–30.

Figure 1: Current and Colonial Boundaries



The figure plots the historical boundaries of direct and indirect British rule, along with the state administrative boundaries, as of 2014. The exact international geographic boundaries have not been verified.

Figure 2: Geography of Project Announcements

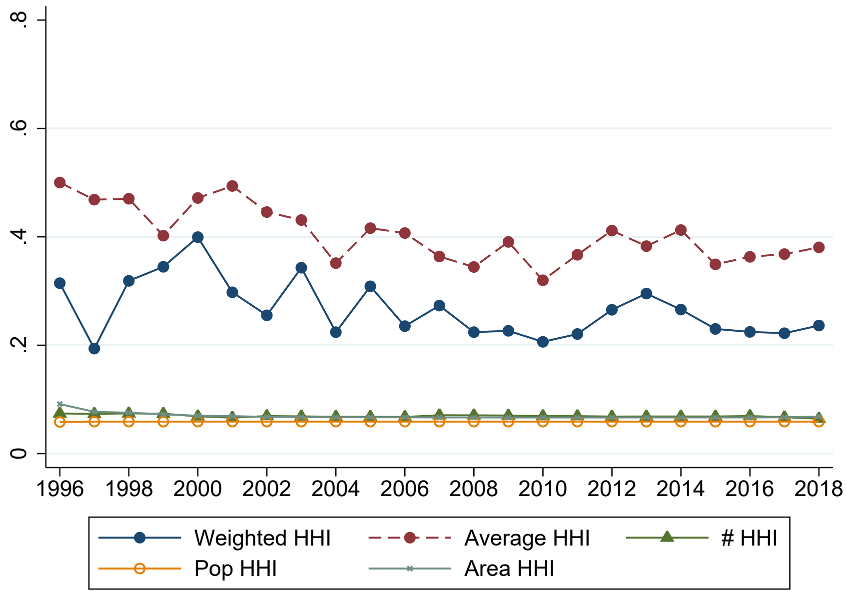


(a) Total Amount (in ₹ bn)

(b) # Projects

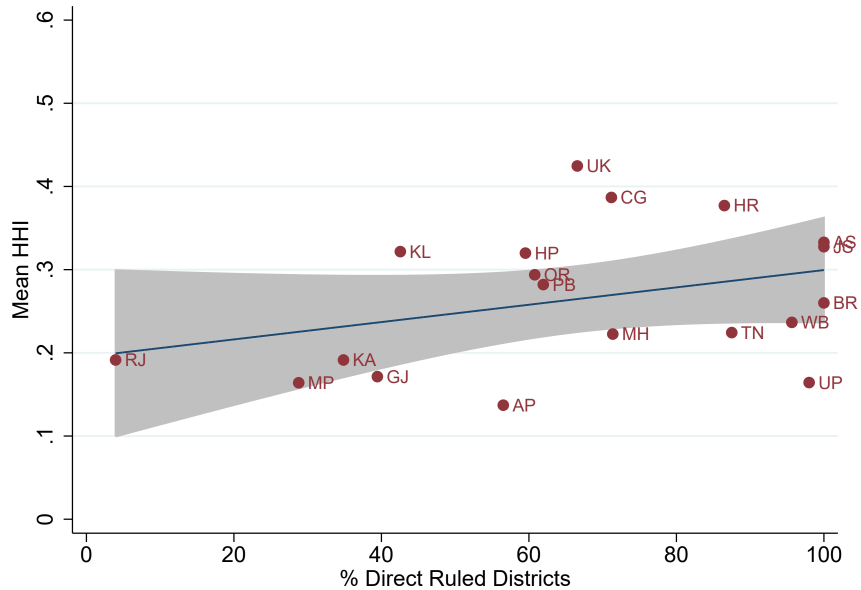
This figure plots the heatmap of total amount (in ₹ billions) and the number of projects announced in the 19 states in our sample between fiscal year 1996 and 2018. Total amount and number of projects are calculated as the total amount and total number of projects announced in each district. The exact international geographic boundaries have not been verified.

Figure 3: Geographic Concentration of Investment



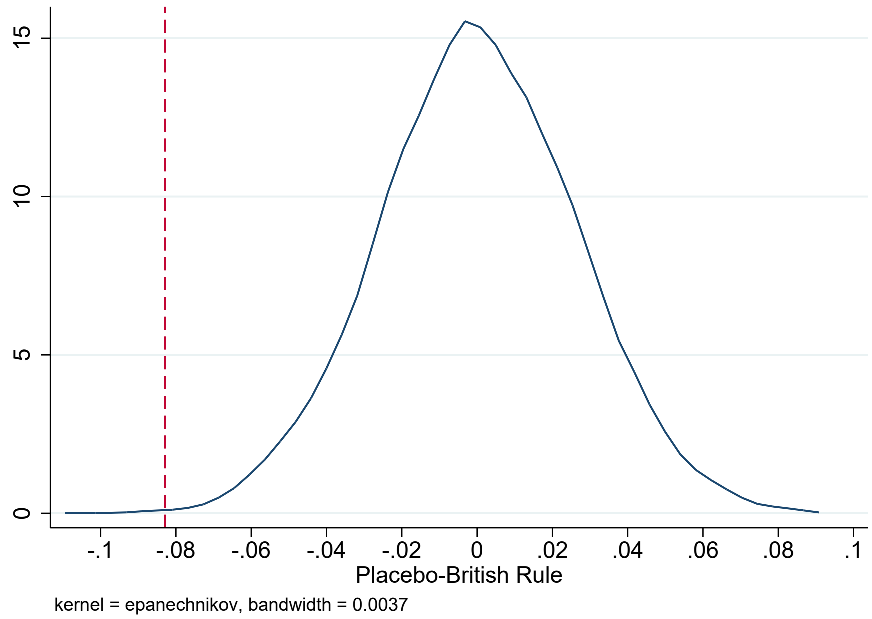
The figure plots a time-varying measure of state-level average and weighted Herfindahl-Hirschman Index (HHI). HHI is computed adding the squared share of investment in a district relative to the state in a year. The red line plots the simple average HHI for a year across all states. The blue line plots the weighted HHI where the HHI of each state is weighted by the share of investment in the state relative to the country. Alongside are three measures of benchmark HHI. # HHI assumes equal investment across all districts and is the simple average of $\frac{1}{N}$, where N denotes the number of districts in a state. State-level Area HHI is computed assuming the investment in each district relative to the state is proportional to the area of the district relative to the state. The aggregate Area HHI is computed by weighing each state by the share of total area in the state. Pop HHI is computed assuming the investment in each district relative to the state is proportional to the population of the district relative to the state population. The aggregate Pop HHI is computed by weighing each state by the share of total population in the state.

Figure 4: Geographic Concentration of Investment and Direct Rule



The figure plots the scatterplot of average state level investment HHI and the percent of direct ruled districts within a state and the best fit line between the two measures. The unit of observation is state. State-level average HHI is computed using data on project announcements from fiscal year 1996 till 2018.

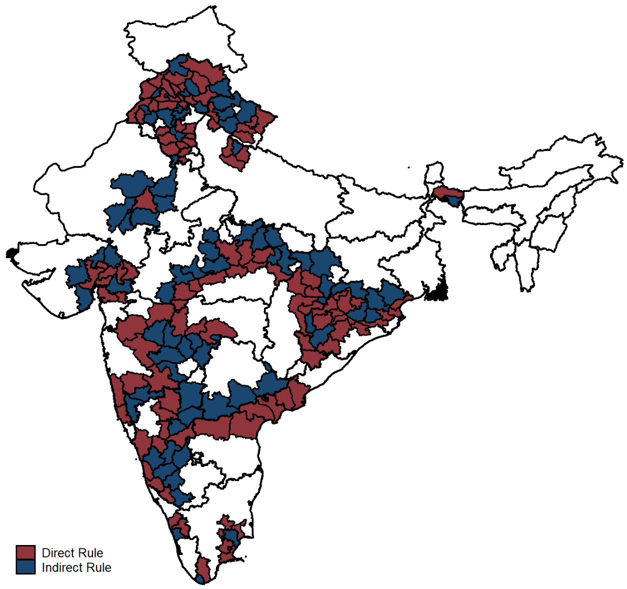
Figure 5: Placebo Test: Randomization of the Direct British Rule



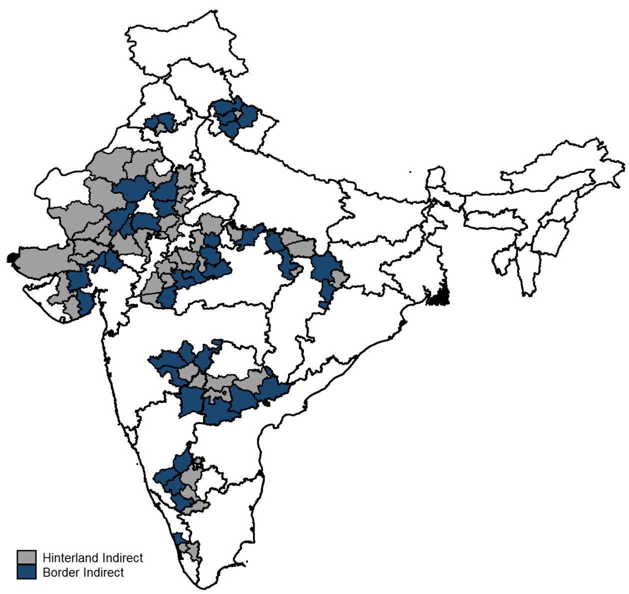
Min	p1	p5	p25	p50	p75	p95	p99	Max	Mean	SD
-0.1058	-0.0604	-0.0432	-0.0174	0.0001	0.0180	0.0436	0.0620	0.0870	0.0003	0.0262

The figure plots the kernel density of the point estimates of $Placebo - Direct - Rule_j$ obtained from the 10,000 Monte-Carlo simulations. A placebo direct rule binary variable is generated for each district. Every district has a probability of 0.6370 to be assigned as directly ruled district. We generate the $Placebo - Direct - Rule_j$ variable 10,000 times and estimate the baseline specification, equation 1, using the generated placebo direct rule variable each time. The table gives the numbers associated with the distribution of the estimates plotted in the figure. The dashed red line shows the point estimate from column (5) of Table 3. There are 0.1% of points to the left of the red-dashed line.

Figure 6: Sample for Local Identification Approach



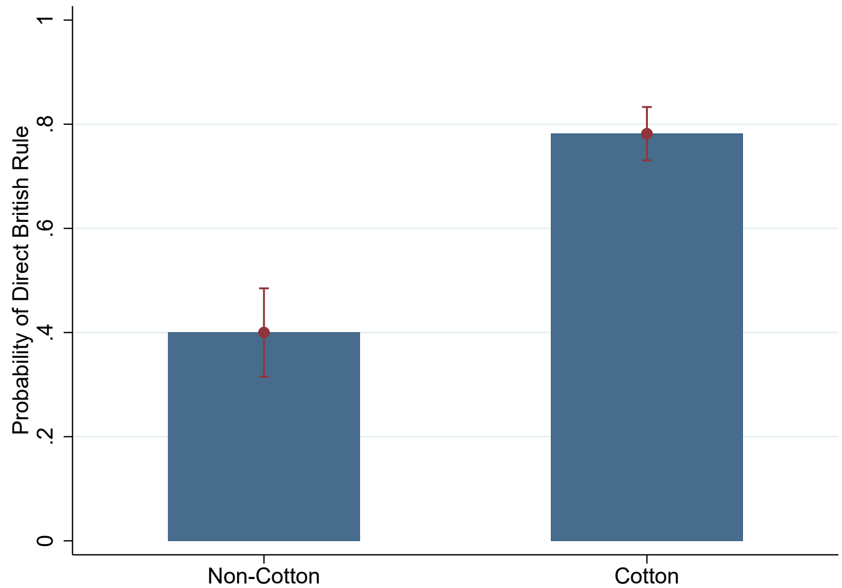
(a) Sample of Bordering Districts



(b) Sample of Hinterland Districts

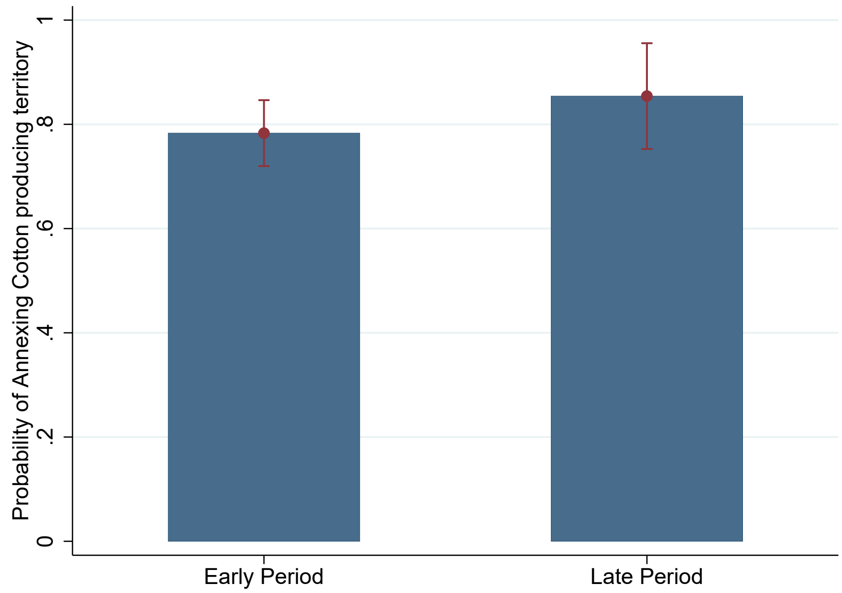
The Figure 6a plots the sample of direct and indirect ruled districts bordering each other within a state, along with the state administrative boundaries, as of 2014. The Figure 6b plots the sample of indirect ruled districts bordering each other within a state, along with the state administrative boundaries, as of 2014. The districts in Figure 6b are chosen based on indirect ruled districts in Figure 6a and the indirect ruled districts immediately bordering them. The exact international geographic boundaries have not been verified.

Figure 7: Probability of Direct British Rule Based on Cotton Production



The figure plots the probability that a district is under direct British rule. The bar graph denotes the likelihood that a district is under direct British rule for non-cotton producing and cotton producing districts. Standard errors are denoted by the red error-bars.

Figure 8: Probability of Cotton Producing Territory under Direct Rule by Timing of Annexation



The figure plots the probability that a district under direct British rule is a cotton producing district. The bar graph denotes the likelihood that a district under direct British rule is a cotton producing territory for early and late annexation periods. Early period refers to annexation before 1800, and late period refers to annexation after 1800. Standard errors are denoted by the red error-bars.

Table 1: What Predicts Direct British Rule?

Dep Var: Direct Rule (=1)	(1)	(2)	(3)	(4)	(5)
Altitude (MSL)	0.0002				0.0002
	(0.0003)				(0.0002)
Coast (=1)	0.1820				0.1720
	(0.1176)				(0.1179)
ln(Area)	-0.0637				-0.0692
	(0.0816)				(0.0799)
Slope	-1.0837				2.6706
	(3.6432)				(2.3314)
Rain (cm)	0.0015				0.0012
	(0.0010)				(0.0009)
Max-Temp	0.0061				-0.0010
	(0.0113)				(0.0113)
Min-Temp	0.0126				0.0028
	(0.0104)				(0.0090)
ln(Distance)		0.0396			0.0707
		(0.0611)			(0.0577)
Maratha Ruler			0.2279		0.2449
			(0.1550)		(0.1524)
Muslim Ruler			0.3853***		0.3319**
			(0.1276)		(0.1420)
Prop Muslim				0.2663	-0.1818
				(0.3447)	(0.2848)
Prop Sikhs				0.6377	-0.2291
				(1.0841)	(0.9907)
Prop Lower Caste				0.5613	0.5439
				(0.3940)	(0.3518)
Prop Elites				-0.3153	-0.1544
				(0.6895)	(0.6948)
Constant	0.5330	0.4275	0.4336***	0.5111***	0.3777
	(0.8825)	(0.3253)	(0.0933)	(0.1445)	(0.9042)
# Obs	294	294	294	294	294
R^2	0.0814	0.0042	0.1257	0.0293	0.1939

The table reports the differences in key observables between direct and direct ruled districts. Column (1) uses geographic covariates such as altitude of the district, coastal dummy indicating if the district is along the coast or not, natural logarithm of district area, average slope. average rainfall, maximum and minimum temperature. Column (2) uses an administrative covariate of the natural logarithm of the distance of the district from the state capital. Column (3) uses two binary variable. Maratha ruler takes a value of 1 if the precolonial ruler belonged to the Maratha empire, the largest empire before the entry of British in India. Muslim ruler takes a value of 1 if the precolonial ruler was of Muslim faith and 0 otherwise. Column (4) includes proportion of Muslim, Sikh, lower caste and elites as per 1911 Indian survey. Elites are defined based on castes that were coded as politically dominant, landowning, or military in the provincial volumes of the 1911 census. Standard errors reported in parentheses are clustered by native states. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Investment Concentration and State Characteristics

Dep Var: HHI	(1)	(2)	(3)	(4)	(5)	(6)
% Direct Rule	0.1213*	0.1463**	0.1227*	0.1522*	0.1514*	0.1369*
	(0.0661)	(0.0531)	(0.0685)	(0.0793)	(0.0813)	(0.0778)
# Districts		-0.0182***	-0.0188***	-0.0180***	-0.0194***	-0.0183***
		(0.0038)	(0.0039)	(0.0042)	(0.0047)	(0.0053)
Area per District			-0.0582	-0.1081	-0.1354	-0.1090
			(0.1128)	(0.1402)	(0.1391)	(0.1541)
Population Density				-0.8775	-1.2064	-0.8094
				(0.8750)	(0.9068)	(1.0235)
GDP per capita					-0.1159	0.0302
					(0.1124)	(0.2965)
% Urban						-0.0087
						(0.0144)
# Obs	19	19	19	19	19	19
R^2	0.1269	0.5086	0.5227	0.5422	0.5753	0.5933

This table presents the estimates of regressions of the average state-level HHI on key state-level explanatory variables. The key explanatory variables include percent of British districts in a state, number of districts in a state, area per district, population density, GDP per capita, and urban population percent. Average values of these variables between 1996 and 2018 are used as explanatory variables. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Investment and Direct British Rule

Dep Var: ln(Project Size)	(1)	(2)	(3)	(4)	(5)
Direct Rule (=1)	-0.1755**	-0.1130***	-0.1146**	-0.0864**	-0.0881***
	(0.0836)	(0.0416)	(0.0526)	(0.0348)	(0.0326)
	[0.0548]***	[0.0356]***	[0.0371]***	[0.0332]***	[0.0331]***
State FE	Yes	Yes	Yes	Yes	
Firm FE		Yes	Yes		
Qtr × Year FE			Yes	Yes	Yes
Firm × Year FE				Yes	Yes
State × Year FE					Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes
# Obs	28,820	28,820	28,820	28,820	28,820
R^2	0.0303	0.5067	0.5465	0.7088	0.7160

The table reports the results of regressing the natural logarithm of project cost on the direct British rule dummy, per the following specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,y} + \theta_{s(j\in s),y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Direct Rule (=1) is a binary variable that takes a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. $\theta_{i,y}$ and $\theta_{s(j\in s),y}$ denote firm × year and state × year fixed effects respectively. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement. Conley (1999) standard errors adjusted for spatial dependence within 100 km are reported in square brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Balanced Panel Analysis: Investment and Direct British Rule

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Announce=1	$I_{j,t}$	$I_{j,t} \| I_{j,t} > 0$	# Projects$_{j,t}$	# Projects$_{j,t} \| \# > 0$	$\frac{I_{j,t}}{\sum_{j \in s} I_{j,t}}$	$\frac{\text{Projects}_{j,t}}{\sum_{j \in s} \text{Projects}_{j,t}}$
Direct Rule (=1)	-0.2534*	-16174.5813**	-28350.1337**	-4.1791**	-6.8549**	-7.0724**	-7.1912*
	(0.1346)	(7910.8774)	(11777.6353)	(2.0257)	(2.7350)	(2.9954)	(3.7650)
State × Qtr × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Average	0.6453	19861.4928	37851.9045	4.4693	7.7296	8.7459	8.8529
	(0.0931)	(5367.5117)	(7901.4838)	(1.3584)	(1.8328)	(2.0333)	(2.5316)
# Obs	35,256	35,256	17,052	35,256	19,050	35,256	35,256
R^2	0.1854	0.2363	0.3115	0.1800	0.1621	0.0500	0.1070

The table reports the results of regressing total project cost and total number of projects and their shares on the direct British rule dummy as per the following specification:

$$I_{j,t} = \beta \cdot \text{Direct Rule}_j + \theta_{s(j \in s),t} + \varepsilon_{j,t}$$

where $Y_{j,t}$ denotes total investment or number of projects in district j, located in state s ($j \in s$) at time t. Direct Rule (=1) is a binary variable taking a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. $\theta_{s(j \in s),t}$ denotes state × quarter × year fixed effects. We aggregate the firm-location-time investment dataset at location-time level, using total investment and total number of projects announced. We create a balanced panel by assigning the value of 0 to a location-time if no projects are announced in the location-time. The dependent variable in column (1) takes a value of 0 if no project is announced in that location-time and 1 otherwise. Column (2) uses total investment in district j as the dependent variable. Column (3) uses total investment in district j, conditional on any investment as the dependent variable. Column (4) uses total number of projects announced in district j. Column (5) uses total number of projects announced, conditional on any announcements in district j. Column (6) uses the share of investment in district j relative to investment across all districts in the state. Column (7) uses the share of the number of projects relative to the total number of projects announced in the state as the dependent variable. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Local Identification Approach: Sample of Bordering Districts

Dep Var: ln(Project Size)	(1)	(2)	(3)
Direct Rule (=1)	-0.0974**	-0.1090**	-0.1084**
	(0.0469)	(0.0488)	(0.0457)
Qtr × Year FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	
District-Pair × Year FE	Yes	Yes	
Firm × District-Pair FE		Yes	
Firm × District-Pair × Year FE			Yes
Lat/Long	Yes	Yes	Yes
# Obs	11,947	11,947	11,947
R^2	0.7856	0.7940	0.7944

The table reports the results of regressing the natural logarithm of project cost on the direct British rule dummy using a sample of bordering districts within a state as shown in Figure 6a, per the following specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,p(j\in p)y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Direct Rule (=1) is a binary variable that takes a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. $\theta_{i,p(j\in p)y}$ denotes the firm × district-pair × year fixed effect. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are multiway clustered by native states, district-pair, and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Local Identification Approach: Sample of Hinterland Districts

Dep Var: ln(Project Size)	(1)	(2)	(3)
Hinterland (=1)	0.0382	0.0353	0.0355
	(0.0549)	(0.0391)	(0.0353)
Qtr × Year FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	
District-Pair × Year FE	Yes	Yes	
Firm × District-Pair FE		Yes	
Firm × District-Pair × Year FE			Yes
Lat/Long	Yes	Yes	Yes
# Obs	4,953	4,953	4,953
R^2	0.8340	0.8431	0.8432

The table reports the results of regressing the natural logarithm of project cost on the hinterland dummy using a sample of bordering districts, both indirect ruled, within states as shown in Figure 6b, per the following specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Hinterland}_j + \theta_{i,p(j \in p)y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Hinterland (=1) is a binary variable that takes a value of 1 if the indirect ruled district is in the hinterland of the sample of direct ruled districts shown in Figure 6a, and 0 otherwise. $\theta_{i,p(j \in p)y}$ denotes the firm × district-pair × year fixed effect. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are multi-way clustered by native states, district-pair, and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Likelihood of Project Abandonment

Dep Var: Project Abandonment	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
Direct Rule (=1)	0.0094	0.0284**	0.0256*	0.0257*	0.0248*
	(0.0121)	(0.0128)	(0.0148)	(0.0153)	(0.0137)
Qtr × Year FE	Yes	Yes	Yes	Yes	Yes
Firm × District-Pair × Year FE	Yes	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes
# Obs	10,938	10,938	10,938	10,938	10,938
R^2	0.5881	0.6125	0.6117	0.6153	0.6231

The table reports the probability of abandonment of project announced in direct ruled districts relative to the projects announced in indirect ruled districts. We identify the status of the project each year, until five years of its announcement, i.e., for a project announced in year x we examine the probability of its abandonment from year $x+1$ until year $x+5$. Our local identification approach includes firm × district-pair × year fixed effects, and quarter × year fixed effects. Lat/Long indicates controls for latitude and longitude of the district. Standard errors are estimated by multiway clustering at the native states, district-pair, and time of project announcement level.

Table 8: Instrumental Variable Regression: Death of Ruler without Natural Heir

Dep Var: ln(Project Size)	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	2SLS		Falsification	
			Second Stage	First Stage		
Direct Rule (=1)	-0.2236***		-0.2239**			
	(0.0596)		(0.0960)			
Rule Death, No Heir, Lapse (=1)		-0.1475*		0.6589***		
		(0.0759)		(0.1225)		
Rule Death, No Heir, No Lapse (=1)					-0.0098	
					(0.0580)	
Rule Death, Yes Heir, Lapse (=1)						0.0072
						(0.0758)
Qtr × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes	Yes
# Obs	10,293	10,293	10,293	10,293	8,034	8,129
R^2	0.763	0.7628	0.0016	0.8646	0.763	0.7626
KP LM Statistic			5.9527**			
KP Wald F-Statistic			28.9431			

The table reports the results of regressing the natural logarithm of project cost on Direct Rule (=1) as per the following 2SLS specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{s(j \in s),y} + \theta_{i,y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

$$\text{Direct Rule}_j = \gamma \cdot \text{Ruler Death-No Heir}_j + \theta_{s(j \in s),y} + \theta_{i,y} + \delta_3 \cdot \text{Latitude}_j + \delta_4 \cdot \text{Longitude}_j + \theta_t + \epsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s $(j \in s)$ at time t during year y. Direct Rule (=1) is a binary variable taking a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. Ruler Death, No Heir (=1) takes a value of 1 if the incumbent Indian ruler dies without a natural heir after 1847 and before 1857. Ruler Death, No Heir, No Lapse (=1) takes a value of 1 if the incumbent Indian ruler dies without a natural heir when the Doctrine of Lapse was not valid – before 1848 and after 1858. Ruler Death, Yes Heir, Lapse (=1) takes a value of 1 if the incumbent ruler dies after 1847 and before 1857 and has a natural heir. The sample for columns 1-4 consists of all districts that were consolidated under direct British rule after 1847 and districts that remained under indirect rule. The sample for columns 5 and 6 include districts that remained under indirect rule. $\theta_{i,y}$ and $\theta_{s(j \in s),y}$ denote firm × year and state × year fixed effects, respectively. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Trade Flows, Cotton, and Timing of Annexation

Dep Var: Trade Flows	(1)	(2)	(3)	(4)	(5)	(6)
Cotton Commodity & Early Annexation	-0.4985***	-0.6209***	-0.6173***	-0.4649***	-0.4417***	-0.4384***
	(0.0974)	(0.1009)	(0.0923)	(0.0359)	(0.0287)	(0.0433)
Early Annexation	0.2020	0.0800	0.0805	0.0985		
	(0.1566)	(0.0657)	(0.0649)	(0.0583)		
Cotton Commodity	0.1365**					
	(0.0518)					
Year FE		Yes				
Commodity FE		Yes				
Export. Prov. × Import. Prov. FE		Yes				
Export. Prov. × Import. Prov. × Year FE			Yes	Yes	Yes	
Commodity × Year FE			Yes	Yes	Yes	Yes
Export. Prov. × Import. Prov. × Commodity FE				Yes	Yes	Yes
Export. Block × Year FE					Yes	
Import. Block × Year FE					Yes	
Export. Block × Import. Block × Year FE						Yes
# Obs	47,447	47,447	47,447	47,447	47,447	47,447
R^2	0.0117	0.2723	0.3125	0.4188	0.5458	0.6569

The table reports the results from regressing the natural logarithm of trade flows for each commodity between trade blocks on the interaction term of early annexation and cotton commodity. Early annexation (=1) takes a value of one for trade blocks that were under direct British rule before 1800, and zero, otherwise. Cotton commodity (=1) takes a value of one for trade flows of cotton and zero for all other commodities. Standard errors reported in parentheses are clustered by exporter province × importer province. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Instrumental Variable Regression: Precolonial Cotton Production

Dep Var: ln(Project Size)	(1) IV	(2) 2SLS Second Stage	(3) 2SLS First Stage	(4) Falsification
Direct Rule (=1)		-0.2272*		
		(0.1270)		
Precolonial Cotton (=1)	-0.2073*		0.9120***	0.0344
	(0.1166)		(0.1326)	(0.0876)
Qtr × Year FE	Yes	Yes	Yes	Yes
Firm × District-Pair × Year FE	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes
# Obs	9,491	9,491	9,491	1,871
R^2	0.7901		0.7786	0.8077
KP LM Statistic			7.3567***	
KP Wald F Statistic			47.3955	

The table reports the results of regressing the natural logarithm of project cost on Direct Rule (=1) as per the following 2SLS specification

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,p(j\in p),y} + \theta_{i,y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$
$$\text{Direct Rule}_j = \gamma \cdot \text{Cotton}_j + \theta_{i,p(j\in p),y} + \theta_{i,y} + \theta_t + \delta_3 \cdot \text{Latitude}_j + \delta_4 \cdot \text{Longitude}_j + \epsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Direct Rule (=1) is a binary variable taking a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. Precolonial cotton (=1) takes a value of 1 if the district produced cotton before annexation. $\theta_{i,p(j\in p),y}$ and θ_t denote firm × district-pair × year and time fixed effects, respectively. Lat/Long indicates controls for latitude and longitude of the district. The falsification test in column 4 uses a sample of bordering indirect ruled district-pairs where one district produced cotton during the precolonial period and other did not. Standard errors reported in parentheses are multiway clustered by native states, district-pair and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Investment and Precolonial Cotton Production

Dep Var: ln(Project Size)	(1)	(2)	(3)
Precolonial Cotton (=1)	-0.3838*	-0.1160	-0.0564
	(0.1879)	(0.1740)	(0.1944)
Early Annexation (=1)		-0.5800***	0.0341
		(0.1938)	(0.1601)
Precolonial Cotton × Early Annexation			-0.6448**
			(0.3073)
Qtr × Year FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes
# Obs	9,465	9,465	9,465
R^2	0.8058	0.8069	0.8070

The table reports the results of regressing present day corporate investment on direct ruled areas based on the timing of annexation. Early annexation (=1) takes a value of one for areas under direct British rule before 1800, and zero, otherwise. Precolonial cotton (=1) for areas that produced cotton. Standard errors are clustered by native states and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Human Capital Formation and Precolonial Cotton Production

Dep Var: 1[Higher Education]	(1)	(2)	(3)	(4)	(5)
Low caste × Early Annexation × Precolonial Cotton	-0.1161**	-0.1235***	-0.1122***	-0.1145***	-0.1132***
	(0.0448)	(0.0391)	(0.0200)	(0.0194)	(0.0220)
Low caste × Early Annexation	0.1254***	0.1337***			
	(0.0262)	(0.0192)			
Low caste × Precolonial Cotton	0.0254	-0.0013	-0.0482**	-0.0460**	-0.0429*
	(0.0254)	(0.0242)	(0.0200)	(0.0194)	(0.0220)
Precolonial Cotton × Early Annexation	0.1026*	0.1092***	0.1189***	0.1278***	
	(0.0506)	(0.0312)	(0.0213)	(0.0215)	
Early Annexation (=1)	-0.1425***				
	(0.0227)				
Low caste (=1)	-0.1622***	-0.1501***			
	(0.0172)	(0.0191)			
Precolonial Cotton (=1)	-0.0192	0.0272	0.0436*	0.0347	
	(0.0265)	(0.0190)	(0.0213)	(0.0215)	
Native State FE		Yes			
Present State FE		Yes			
Native State × Caste FE			Yes	Yes	Yes
Present State × Caste FE			Yes	Yes	Yes
Annexation Year FE				Yes	
District FE					Yes
Lat/Long					Yes
# Obs	41,278	41,278	41,278	41,278	41,278
R^2	0.0422	0.0604	0.0751	0.0767	0.0993

The table reports the results of regressing an indicator for whether the household hold has received higher education on direct ruled areas based on the timing of annexation, cotton production, and caste. Low caste (=1) takes a value of one if the household head belongs to a scheduled tribe, scheduled caste, or other backward caste. Early annexation (=1) takes a value of one for areas under direct British rule before 1800, and zero, otherwise. Precolonial cotton (=1) for areas that produced cotton. Our sample only includes districts that were under direct British rule. Education data comes from the National Demographic and Health Survey 1998/1999 wave. Standard errors are clustered by native states. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Interest in Acquiring Education and Precolonial Cotton Production

Dep Var: 1[No Interest]	(1)	(2)	(3)	(4)	(5)
Low caste × Early Annexation × Precolonial Cotton	0.3846***	0.2720***	0.2643***	0.2714***	0.1855***
	(0.0940)	(0.0695)	(0.0567)	(0.0524)	(0.0512)
Low caste × Early Annexation	-0.3303***	-0.2610***			
	(0.0656)	(0.0632)			
Low caste × Precolonial Cotton	0.0351	0.0211	0.0409	0.0338	-0.0367
	(0.0476)	(0.0300)	(0.0567)	(0.0524)	(0.0512)
Precolonial Cotton × Early Annexation	-0.3323***	0.1340*	0.1309**	0.1311**	
	(0.0779)	(0.0695)	(0.0508)	(0.0512)	
Early Annexation (=1)	0.3067***				
	(0.0330)				
Low caste (=1)	-0.0306	-0.0346			
	(0.0260)	(0.0224)			
Precolonial Cotton (=1)	-0.0044	-0.0740**	-0.0825	-0.0827	
	(0.0256)	(0.0343)	(0.0508)	(0.0512)	
Native State FE		Yes			
Present State FE		Yes			
Native State × Caste FE			Yes	Yes	Yes
Present State × Caste FE			Yes	Yes	Yes
Annexation Year FE				Yes	
District FE					Yes
Lat/Long					Yes
N	14,322	14,322	14,322	14,322	14,322
R^2	0.0027	0.0756	0.0869	0.0882	0.1319

The table presents regression results where the dependent variable is an indicator for whether the primary reason a household member did not attend school was either that education was not considered necessary or there was a lack of interest in studies. Key independent variables include indicators for directly ruled areas based on annexation timing, cotton production, and caste. Alternative reasons for not attending school include factors such as distance from school, the need to work on the family farm or in a family business, high costs, inadequate school facilities for girls, and caregiving responsibilities for siblings. Low caste (=1) takes a value of one if the household head belongs to a scheduled tribe, scheduled caste, or other backwards caste. Early annexation (=1) takes a value of one for areas under direct British rule before 1800, and zero, otherwise. Precolonial cotton (=1) for areas that produced cotton. Education data comes from the National Demographic and Health Survey 1998/1999 wave. Our sample only includes districts that were under direct British rule. Standard errors are clustered by native states. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Internet Appendix for:

"What Explains Geographic Variation in Corporate Investment?"

Appendix A British Rule in India

British colonial rule over the Indian subcontinent began in 1757 and lasted until 1947. After the decline of the Mughals and the Maratha Empire, India was divided into several small states ruled by royal families, referred to as *princely states*. The British East India Company (East India Company) started its rule in India after defeating the erstwhile Nawab of Bengal, in the Battle of Plassey, in 1757. By 1773, the Company was involved in the direct administration of the areas of present day West Bengal, Bangladesh, and Bihar, with the appointment of the first Governor-General, Warren Hastings. Despite that annexation of India was spearheaded by the East India Company, the crown controlled its activities through various means. The Regulating Act of 1773 set up a system whereby the government supervised and regulated the work of East India Company. Pitt's India Act of 1774 provided for the appointment of a Board of Control and gave the government ultimate authority. East India Company's autonomous control of the subcontinent was successively reduced over time.[16] Further, the Charters of 1813 asserted the Crown's undoubted sovereignty over all East India Company territories, while the 1833 Charter Act invested the Board of Control with full authority over East India Company and increased the power of the governor-general. While the subcontinent was initially under the control of the East India Company, there was a formal transfer of power from the Company to the British Crown after the Sepoy Mutiny of 1857, which marked the end of British expansion in India.

The British Empire consolidated its grip on political control across present day India, Bangladesh, Burma, and Pakistan in "British India" under two broad policies of annexation namely Subsidiary Alliance and the Doctrine of Lapse. The *ring-fence* period between 1765 and 1818, was the first wave of British annexation, under the leadership of Lord Wellesley, the British Governor-General of India between 1798 and 1805. In this period, the East India Company sought to maintain a ring fence between its territories and the empires of the Marathas and Sikhs. One method of achieving this was the Subsidiary Alliance. Under the Subsidiary Alliance, the East India Company

[16] Parliament controlled the East India Company by extending its charter 20 years at a time. Those granted in 1793, 1813, 1833, and 1853 successively whittled away the East India Company's commercial rights and trading monopolies.

entered contracts with princely states, providing the latter with the subsidiary militia for protection and necessary defense for payment. In the event of default by the Indian ruler, a part of their territory was surrendered to the British. The states of Hyderabad (1798), Mysore (1799), Awadh (1801), Peshwa (1802), Bhonsle and Scindia (1803), Udaipur, Jodhpur, and Jaipur (1818) were annexed under the Subsidiary Alliance. The second wave of British annexation, *subordinate isolation*, began in 1818, lasting till the Indian Mutiny of 1857. This phase was marked by the policy of Doctrine of Lapse, under the leadership of its chief architect, Lord Dalhousie, the British Governor-General of India between 1848 and 1856. Under the Doctrine of Lapse, any subordinate princely state would automatically lapse to the British East India Company if the ruler died without a natural heir. The states of Satara (1848), Sambalpur (1849), Baghat (1850), Jaipur of Bundelkhand (1849), Udaipur of Rajputana (1852), Jhansi (1853) and Nagpur (1854) were annexed under Doctrine of Lapse.

Power was transferred from the East India Company to the British Crown after the Sepoy Mutiny of 1857.[17] Certain areas that had been annexed by the British were integrated under direct rule, while remaining areas, were granted indirect rule in which authority was delegated to local princes (princely states). Despite the Queen's proclamation of 1858 ruled out any future annexation, the Governor-General did retain the power to interfere in the internal matters of the princely states in case of a "misrule." Ashton (1977) documents active interference by the British colonial power in the internal affairs of princely states, e.g., Lord Curzon, the Governor-General of India from 1899 to 1905 forced fifteen rulers to abdicate, during his tenure on account of misrule.

Princely states were heterogeneous – some consisted of a few villages, while others, over thousands of square miles. Some princely states had maximum administrative powers while other princely states had minimal administrative powers.[18] Iyer (2010) notes that the princely states constituted approximately 45% of the total geographic area of present day India, and 23% of total population in 1911. Additionally, she notes the presence of princely states across India with a high concentration in central and western India. The Foreign Office recognized about 680 Indian princely states in the year 1910.

After Indian Independence in 1947, areas under direct British rule federated. The princely

[17] This is referred as the Government of India Act 1858 – an Act of the British Parliament, which nationalized the East India Company and granted the British monarch supreme authority over India, as well as power and possessions of the East India Company.

[18] Princely states had varying degrees of legal autonomy, ranging from first-class, wherein the state could try criminal cases to third-class states whereby only small civil cases could be adjudicated by the ruler.

states, however, could choose to join the Indian Union or remain independent. Then Prime Minister Jawaharlal Nehru and the Home Minister Sardar Vallabhbhai Patel, integrated all princely states into the Indian Union by 1950. Thereafter, princely rulers no longer maintained sovereignty, but played a major role in post-independence politics. Allen and Dwivedi (1984) document election of several former princes to federal and state-level political offices.

Table A.1: British Annexation - Time and Reason

	Ceded	Conquest	Grant	Lapse	Misrule	Total
Initial Settlement	0	6	3	0	0	9
Ring Fence (1765-1818)	58	114	15	0	3	190
Subordinate Isolation (1819-1856)	5	22	0	27	17	71
Post 1857 Revolt	2	0	0	0	0	2
Direct Ruled	65	142	18	27	20	272
Indirect Ruled						152
Total						424

The table reports the distribution of the number of districts that were invaded in the CapEx sample, over time and by method of annexation. These districts span a total of 17 states in the sample. The modern day districts are defined as of 2014. Time refers to the four periods marked by different policices - (1) the initial settlement period before 1765, (2) the ring fence period from 1765 until 1818, (3) subordinate isolation from 1819 until 1856, and (4) period of no annexation after the Indian Revolt of 1857.

Appendix B Destruction of Cotton Industry

In this section, we emphasize four ways through which the textile industry, specifically the cotton industry, was destroyed in India: (1) protectionism of the British textile industry through tariffs, (2) price fixing through monopsony power, (3) violence against textile producers, and (4) deprivation of new technological innovations in the local industries.

First, the enactment of protectionist policies and tariffs enabled the British to monopolize trade.[19] By 1813, British textiles faced 5% import tax in India, compared to up to 85% for Indian textiles in Britain (Alavi (1982); Nakatomi (2008)).

Second, the British forced the weavers to exclusively sell their goods to them at fixed prices, establishing themselves as a monopsony. Prices were often fixed below costs, causing weavers to fall into poverty (Mukund (1992)). Moreover, local agents, *gomasthas*, monitored production, ensuring that weavers could not privately sell to other buyers. As a result, the textile industry in India deteriorated. The real wages of Indian laborers in 1820 reached 45% of their wages in 1600, substantially reducing their purchasing power (Cypher (2008)). The economic abuse of Indian cotton producers is best portrayed in the March 1921 Young India magazine essay by Mahatma Gandhi – "The labour of these artisans was so cruelly suppressed that they were obliged to cut off their own thumbs in order to avoid imprisonment...In my opinion, such cutting off would be less cruel than the terrorism which resulted in self-mutilation" (Gandhi (1921)).

Third, in addition to economic abuse, there is evidence that laborers experienced physical abuse. Bolts, a British merchant and judge of the mayor's court of Calcutta recorded,

> "With every species of monopoly, therefore every kind of oppression to manufactures of denominations throughout the country has daily increased; insomuch that weavers for daring to sell their goods, and Dallals and Pykars, for having contributed to or connived at such sales, have, by the Company's agents, been frequently seized and imprisoned, confined in irons, fined considerable sums of money, flogged, and deprived, in the most ignominious manner. . . Weavers also, upon their inability to perform such agreements as have been forced from them by the Company's agents ... have had their goods seized, and sold on the spot, to make good

[19] The British have a long history of taking actions to protect its domestic textile industry and eliminate the competitiveness of the Indian textile industry, before the onset of colonization. This included the enactment of several bills. In 1685, a 10% tariff was imposed on the import of Indian goods. In 1690, this tariff was increased to 20%. In 1700, the First Calico Act was passed which banned the import of painted, dyed, stained, and printed fabric. In the same year, the British abolished the export duty on English woolen products, to improve the competitiveness of the British textile industry. In 1707, the tariff on Indian goods was increased to 50%. These acts of protectionism culminated in the Second Calico Act of 1721, which banned the sale of most cotton textiles from India.

> the deficiency: and the winders of raw silk, called Nagaads, have been treated also with such injustice, that instances have been known of their cutting off their thumbs, to prevent their being forced to wind silk" (Bolts (1775)).

Fourth, there is scant evidence that the British encouraged industrialization of cotton mills. Charles Wood, the Secretary for State in India, discouraged the British Government from promoting the use of cotton mills in India to produce equivalent goods that Britain could produce (Gopal (1965)). The British attitude towards the domestic industrial expansion in India was laissez-faire (Avineri (1969)). Sahoo (2015) writes "the financier had not entered the field" indicating lack of supply of capital to India, preventing the development of the Indian cotton mills in the same way the mills of Lancashire had developed.

Overall the attitude of the British towards the Indian cotton textile production is best described by HH Wilson, in his continuation of James Mill's *The History of British India*.

> ". . . had not such prohibitory duties and decrees existed, the mills of Paisley and of Manchester would have been stopped in their outset, and could hardly have been again set in motion, even by the powers of steam. They were created by the sacrifice of the Indian manufactures. Had India been independent, she would have retaliated; would have imposed preventive duties upon British goods, and would thus have preserved her own productive industry from annihilation. This act of self-defence was not permitted her; she was at the mercy of the stranger. British goods were forced upon her without paying any duty; and the foreign manufacturer employed the arm of political injustice to keep down and ultimately strangle a competitor with whom he could not contend on equal terms" (Mill and Wilson (1848)).

Appendix C Data

C.1 Project Announcement - CapEx Database

Table C.1: Project Announcements by Fiscal Year

Year	# Projects	Total Amt (in tn)	Mean Amt (in bn)	Median Amt (in bn)
1996	588	3.71	6.31	0.53
1997	596	4.02	6.75	1.06
1998	367	4.85	13.20	1.22
1999	392	1.52	3.88	0.60
2000	456	1.83	4.01	0.48
2001	1,089	2.49	2.28	0.20
2002	1,033	2.89	2.80	0.22
2003	1,200	2.68	2.23	0.15
2004	1,478	8.13	5.50	0.40
2005	839	6.11	7.29	0.65
2006	1,154	15.10	13.11	0.84
2007	1,677	24.20	14.43	1.00
2008	1,608	22.00	13.69	1.39
2009	1,682	25.00	14.88	2.39
2010	2,040	26.00	12.73	1.65
2011	2,378	23.00	9.68	1.04
2012	1,902	17.50	9.18	1.00
2013	1,218	10.50	8.60	1.08
2014	1,060	6.66	6.29	1.40
2015	1,374	22.30	16.26	2.23
2016	1,605	13.20	8.19	1.45
2017	1,648	19.80	12.00	0.98
2018	1,436	8.93	6.22	1.24
Total	28,820	272.41	9.45	1.00

The table reports the total number of projects, the mean, the median value of projects (in ₹ billion), and the total value of projects (in ₹ trillion) announced across 17 states by all firms between fiscal year 1996 and 2018.

Table C.2: Project Announcements by Project Size

Project Size	# Projects	Total Amt (in tn)	Mean Amt (in bn)	Median Amt (in bn)
<0.3 bn	8,808	0.97	0.11	0.09
0.3-3 bn	11,088	12.60	1.14	0.98
3-10 bn	4,560	24.40	5.35	5.00
>10 bn	4,364	234.00	53.72	22.74
Total	28,820	272.41	9.45	1.00

The table reports the total number of projects, the mean, the median value of projects (in ₹ billion), and the total value of projects (in ₹ trillion) announced across 17 states by project size between fiscal year 1996 and 2018.

Table C.3: Project Announcements by Project Type

Project Type	# Projects	Total Amt (in tn)	Mean Amt (in bn)	Median Amt (in bn)
New Unit	20,510	208.00	10.16	0.91
Renovation & Modernisation	3,318	12.50	3.76	0.43
Substantial Expansion	4,992	51.50	10.32	2.07
Total	28,820	272.41	9.45	1.00

The table reports the total number of projects, the mean, the median value of projects (in ₹ billion), and the total value of projects (in ₹ trillion) announced across 17 states by project type between fiscal year 1996 and 2018.

Appendix D Robustness

Table D.1: Balanced Panel Analysis: Investment and Direct British Rule

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln(1+I_{j,t})$	$\ln(I_{j,t})$	ln(1+# Projects$_{j,t}$)	ln(# Projects$_{j,t}$)	$I_{j,t}$	# Projects$_{j,t}$
Direct Rule (=1)	-2.6273*	-1.6590**	-0.7431**	-1.1878**	-1.7940**	-2.6497***
	(1.3827)	(0.6903)	(0.3693)	(0.4745)	(0.7804)	(0.8618)
State × Qtr × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
# Obs	35,256	17,052	35,256	19,050	35,256	35,256
R^2	0.2008	0.2655	0.2399	0.1894		
Model	OLS	OLS	OLS	OLS	Poisson	Poisson

The table reports the results from the following regression specification:

$$y_{j,t} = \beta \cdot \text{Direct Rule}_j + \theta_{s(j \in s),t} + \varepsilon_{j,t}$$

where $y_{j,t}$ denotes the natural logarithm of 1+total investment in district j, located in state s ($j \in s$) at time t in column 1, the natural logarithm of total investment in column 2, the natural logarithm of 1+number of projects in column 3, and the natural logarithm of the number of projects in column 4. Columns 5 and 6 use the total investment and number of projects as the dependent variable. Columns 1-4 report estimates using an OLS model. Columns 5 and 6 report estimates using a Poisson pseudo-maximum likelihood model. Direct Rule (=1) is a binary variable taking a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. $\theta_{s(j \in s),t}$ denotes state × quarter × year fixed effects. We aggregate the firm-location-time investment dataset at location-time level, using total investment announced. We create a balanced panel by assigning the value of 0 to a location-time if no projects are announced in the location-time. Column (1) includes all district-year observations and column (2) includes a district-year observation only if there was at least one project announced there. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2: Investment and Direct British Rule: Controlling for District Level Geographic Characteristics

Dep Var: ln(Project Size)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Direct Rule (=1)	-0.0692**	-0.0916***	-0.1045***	-0.1025***	-0.0782**	-0.0926**	-0.0846**	-0.0856**
	(0.0317)	(0.0344)	(0.0340)	(0.0348)	(0.0346)	(0.0355)	(0.0323)	(0.0376)
Altitude (MSL)	-0.0001							-0.0001
	(0.0001)							(0.0001)
Coast (=1)		0.0210						0.0050
		(0.0470)						(0.0493)
ln(Area)			0.0296					0.0423
			(0.0274)					(0.0293)
Slope				0.6204				0.7599
				(0.7148)				(1.1256)
Rain (cm)					-0.0007**			-0.0006*
					(0.0003)			(0.0004)
Max-Temp						0.0042		-0.0004
						(0.0031)		(0.0033)
Min-Temp						0.0026		-0.0008
						(0.0041)		(0.0038)
ln(Distance)							-0.0217***	-0.0437**
							(0.0065)	(0.0186)
Qtr × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Obs	27,223	28,820	24,470	24,470	28,820	23,499	28,820	21,181
R^2	0.7195	0.7160	0.7318	0.7317	0.7161	0.7302	0.7161	0.7380

The table reports the results of regressing the natural logarithm of project cost on the direct British rule dummy after controlling for district level characteristics, per the following specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \Theta X_j + \theta_{i,y} + \theta_{s(j\in s),y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Direct Rule (=1) is a binary variable that takes a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. X_j refers to a vector of non-time varying geographic characteristics including altitude (MSL), coastal dummy, natural logarithm of geographic area, slope of land, rainfall (in cm), maximum and minimum temperature, and the natural logarithm of the distance of the district from state capital. $\theta_{i,y}$ and $\theta_{s(j\in s),y}$ denote firm × year and state × year fixed effects respectively. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Investment and Direct British Rule: Controlling for Precolonial Characteristics

Dep Var: ln(Project Size)	(1)	(2)	(3)
Direct Rule (=1)	-0.0831**	-0.1021***	-0.0932**
	(0.0368)	(0.0339)	(0.0356)
Maratha Ruler	-0.0330		-0.0330
	(0.0398)		(0.0415)
Muslim Ruler	-0.0212		-0.0375
	(0.0151)		(0.0302)
Prop Muslim		-0.0249	-0.0083
		(0.2007)	(0.2025)
Prop Sikhs		-0.3338	-0.3783*
		(0.2109)	(0.2225)
Prop Lower Caste		0.0646	0.0634
		(0.1195)	(0.1195)
Prop Elites		-0.0742	-0.0541
		(0.2140)	(0.2145)
Qtr × Year FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes
# Obs	19,800	19,800	19,800
R^2	0.7305	0.7305	0.7305

The table reports the results of regressing the natural logarithm of project cost on the direct British rule dummy after controlling for district level characteristics, per the following specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \Theta X_j + \theta_{i,y} + \theta_{s(j \in s),y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Direct Rule (=1) is a binary variable that takes a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. X_j refers to a vector of non-time varying characteristics associated with the demographics of the district. Maratha ruler takes a value of 1 if the precolonial ruler belonged to the Maratha empire, the largest empire before the entry of British in India. Muslim ruler takes a value of 1 if the precolonial ruler was of Muslim faith and 0 otherwise. We include the proportion of Muslim, Sikh, lower caste and elites as per 1911 Indian survey. Elites are defined based on castes coded as politically dominant, landowning, or military in the provincial volumes of the 1911 census. $\theta_{i,y}$ and $\theta_{s(j \in s),y}$ denote firm × year and state × year fixed effects respectively. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4: Instrumental Variable Regression: Precolonial Cotton Production

Dep Var: ln(Project Size)	(1)	(2)	(3)
	IV	2SLS	
		Second Stage	First Stage
Direct Rule (=1)		-0.3119***	
		(0.1078)	
Precolonial Cotton (=1)	-0.1293***		0.4144***
	(0.0372)		(0.1119)
Qtr × Year FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes
# Obs	19,800	19,800	19,800
R^2	0.7305		0.6738
KP LM Statistic			10.0881***
KP Wald F Statistic			13.7112

The table reports the results of regressing the natural logarithm of project cost on Direct Rule (=1) as per the following 2SLS specification

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,y} + \theta_{s,y} + \theta_t + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

$$\text{Direct Rule}_j = \gamma \cdot \text{Cotton}_j + \theta_{i,p(j\in p),y} + \theta_{i,y} + \theta_{s,y} + \theta_t + \delta_3 \cdot \text{Latitude}_j + \delta_4 \cdot \text{Longitude}_j + \epsilon_{i,j,t}$$

where i denotes firm announcing a project of size $Y_{i,j,t}$ in district j, located in state s ($j \in s$) at time t during year y. Direct Rule (=1) is a binary variable taking a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. Precolonial cotton (=1) takes a value of 1 if the district produced cotton before annexation. $\theta_{i,y}$, $\theta_{s,y}$ and θ_t denote firm × year, state × year, and time fixed effects, respectively. Lat/Long indicates controls for latitude and longitude of the district. Standard errors reported in parentheses are two-way clustered by native states and time of project announcement for columns (1)-(3) and multiway clustered by native states, district-pair and time of project announcement for columns (4)-(6). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: Alternative Measures of Human Capital Formation and Precolonial Cotton Production

	(1)	(2)	(3)	(4)	(5)
Dep Var:	1[No Education]	1[Primary Edu.]	1[Secondary Edu.]	1[Literate]	# Educ. Year
Low caste × Early Annexation × Precolonial Cotton	0.6912***	-0.2613***	-0.3174***	-0.6927***	-5.0703***
	(0.0518)	(0.0038)	(0.0596)	(0.0442)	(0.3787)
Low caste × Precolonial Cotton	0.0701	-0.0323***	0.0057	-0.0810*	-0.9284**
	(0.0518)	(0.0038)	(0.0596)	(0.0442)	(0.3787)
Native State × Caste FE	Yes	Yes	Yes	Yes	Yes
Present State × Caste FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes
# Obs	41,175	41,175	41,175	41,175	41,175
R^2	0.1172	0.0439	0.0558	0.1212	0.1710

The table reports regression results for five dependent variables: an indicator for whether the household head has not received any formal education (column 1), whether the highest educational level attained for the household head is primary education (column 2), whether the highest educational level attained for the household head is secondary education (column 3), an indicator for whether the household head can read and write (column 4), and the number of years of education completed by the household head (column 5). The key independent variables include indicators for directly ruled areas based on annexation timing, cotton production, and caste. Low caste (=1) takes a value of one if the household head belongs to a scheduled tribe, scheduled caste, or other backwards caste. Early annexation (=1) takes a value of one for areas under direct British rule before 1800, and zero, otherwise. Precolonial cotton (=1) for areas that produced cotton. Education data comes from the National Demographic and Health Survey 1998/1999 wave. Our sample only includes districts that were under direct British rule. Standard errors are clustered by native states. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix E Spatial Autocorrelation

We calculate Moran's I statistic to test for global spatial autocorrelation. This provides a gauge of how randomly distributed the direct rule variable is in our analysis. The Moran I statistic measures the interdependency between different regions and ranges from -1 to 1. Under random distribution, the statistic approaches zero, asymptotically. A statistic value above zero reflects positive spatial autocorrelation between districts i and j. A statistic value below 0 reflects negative spatial autocorrelation between districts i and j.

Moran's I statistic is computed as follows:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{E.1}$$

where n is the number of districts, z_i is an indicator for direct rule in district i – z_i is standardized, w_{ij} denotes the ij^{th} element of a row-standardized weight matrix which uses distance between i and j as weights.

E.1 Construction of Weight Matrix

In the weight matrix, diagonal elements are 0 (the distance between a region and itself is 0).

$$W = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,m-1} & w_{1,m} \\ w_{2,1} & 0 & \cdots & w_{2,m-1} & w_{2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{m-1,1} & w_{m-1,2} & \cdots & 0 & w_{m-1,m} \\ w_{m,1} & w_{m,2} & \cdots & w_{m,m-1} & 0 \end{bmatrix} \tag{E.2}$$

In our analysis, we use two types of spatial weight matrices: inverse weighting and uniform weighting.

E.1.1 Inverse Weights

Inverse weights between districts i and j are constructed in the following way.

$$w_{ij} = \begin{cases} \frac{d_i j^{-\delta}}{\sum_{j=1}^{n} d_i j^{-\delta}} & d_{ij} < d, i \neq j, \delta > 0 \\ 0 & otherwise \end{cases} \tag{E.3}$$

where d is a distance threshold, which we vary from 10 miles to 200 miles to study the robustness of the statistic. δ is a decay parameter for distance. It takes the value 1.

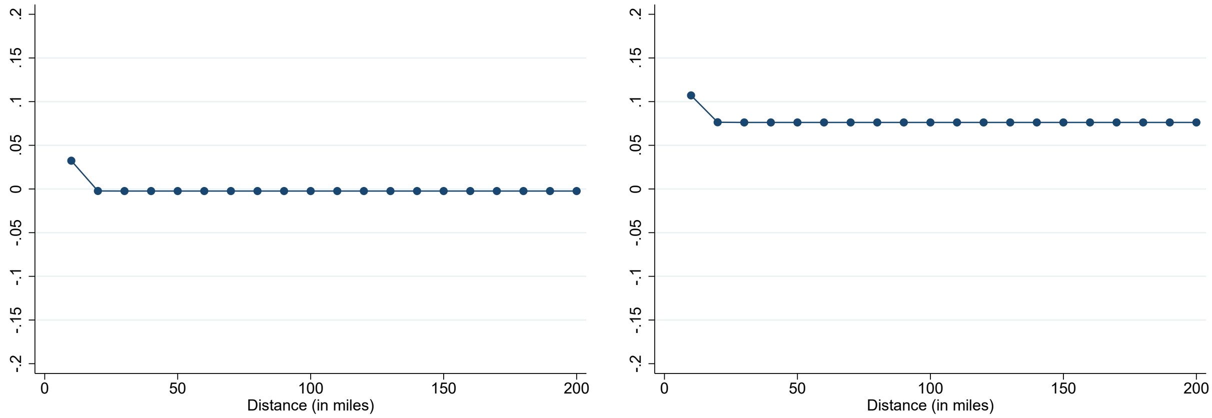
E.1.2 Uniform Weights

Uniform weights between i and j are constructed in the following way: $w_{ij} =$

$$\begin{cases} \frac{1_{d_{ij}<d}}{\sum_{j=1}^{n} 1_{d_{ij}<d}} & d_{ij} < d, i \neq j \\ 0 & otherwise \end{cases}$$

where $1_{d_{ij}<d}$ is an indicator function that takes the value 1 if the distance between districts i and j is below the threshold distance d and 0, otherwise.

Figure E.1: Spatial Autocorrelation - Moran I statistic



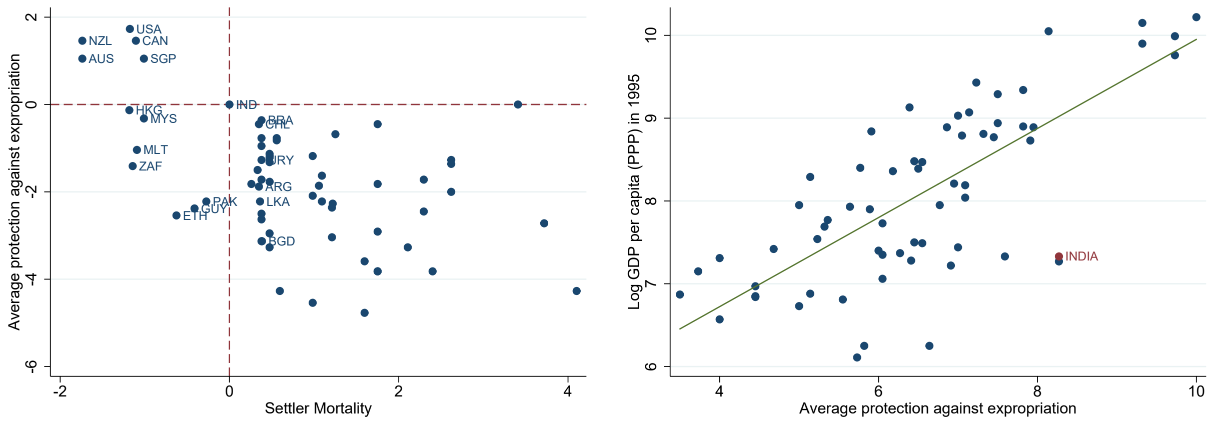
(a) Uniformly Weighted

(b) Inverse Distance Weights

The figure plots the Moran I statistic for the Direct British Rule variable. The Moran I is computed for different distance thresholds from 10 miles to 200 miles. Figure E.1a uses the uniform weight between two points if they are within the distance limit and 0 otherwise. The weight takes a value of 1 if the two points are within the threshold and 0 otherwise. Figure E.1b uses the weights calculated as the inverse of the distance between two points if they are within the distance limit and 0 otherwise. Weights are standardized such that they add up to 1 within a row.

Appendix F Alternative Explanations:

Figure F.1: Protection Against Expropriation - Acemoglu, Johnson and Robinson (2001)



(a) Protection Against Expropriation & Settler Mortality

(b) Protection Against Expropriation & Economic Prosperity

The figure replicates the two key results presented in Acemoglu, Johnson and Robinson (2001) for their sample. Panel a shows the scatterplot of average protection against expropriation with settler mortality. We standardize the data such that the two measures take a value of zero for India. Panel b shows the scatterplot of log GDP per capita (PPP) in 1995 against the average protection against expropriation with the best-fit line. The red scatter dot in panel b denotes India.

Table F.1: Court Delays and Direct British Rule

Dep Var: ln(Mean Delay)	(1)	(2)	(3)
Direct Rule (=1)	-0.0025	-0.0025	-0.0025
	(0.0245)	(0.0245)	(0.0243)
Statute × Year FE	Yes	Yes	
District-Pair × Year FE	Yes	Yes	
Statute × District-Pair FE		Yes	
Statute × District-Pair × Year FE			Yes
Lat/Long	Yes	Yes	Yes
# Obs	180,580	180,580	180,580
R^2	0.6155	0.7204	0.8077

The table reports the results of regressing the natural logarithm of district level mean delay in cases by statutes on the direct British rule dummy using a sample of bordering districts within states as shown in Figure 6a, per the following specification:

$$ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_i + \theta_{j,p(i\in p),t} + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \varepsilon_{i,j,t}$$

where i denotes a district with a mean duration of $Y_{i,j,t}$ for statute j during year t. Direct Rule (=1) is a binary variable that takes a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. $\theta_{j,p(i\in p),t}$ denotes the statute × district-pair × year fixed effect. Lat/Long indicates controls for latitude and longitude of the district. Data on court records comes from Ash et al. (2021). The dataset draws from 80 million court records spanning 2010-2018, sourced from the Indian e-Courts platform – an online database implemented by the Government of India, detailing information on the universe of India's 7,000+ district and subordinate trial courts and over 80,000 judges. The Indian e-Courts database provides comprehensive coverage of India's lower judiciary, consisting of District and Session courts. Our focus of the data is on the delay of a case. Delay is the time that a court takes to resolve a case. It is measured as the number of days between the initial filing date and the decision date. At the district-level, the median duration of a case is 241 days. There is a large heterogeneity in case delay. The 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of case duration are 28 days and 746 days, respectively. Moreover, the total number of cases also exhibit a large heterogeneity. Standard errors reported in parentheses are two-way clustered by native states and district-pair. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Public Goods: We obtain data on public goods from the Population Census of 1991, 2001 and 2011, as reported in the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG). This dataset provides detailed information on public goods. It reports the number of schools at various levels, including primary, middle, lower secondary, and senior secondary at the village and town level, as well as the prevalence of rural electrification. We aggregate this data at the district-year level to compute the number of schools per 1,000 people and the fraction of villages with electricity connection.

Table F.2: Provision of Public Goods: Schools & Electricity

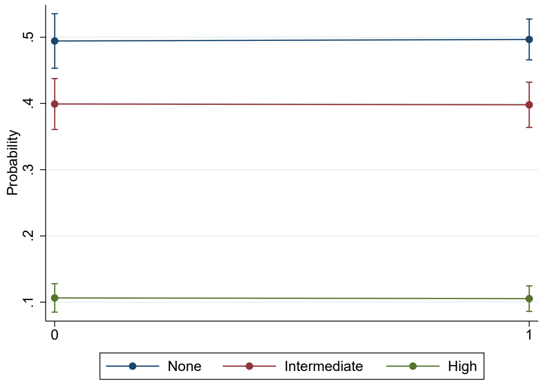
	(1)	(2)	(3)	(4)	(5)
	Total School	Primary School	Middle School	High School	Electricity
Direct Rule (=1)	-0.0844	-0.0536	-0.0248	-0.0059	0.0114
	(0.0887)	(0.0601)	(0.0187)	(0.0128)	(0.0115)
District-Pair × Year FE	Yes	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes
# Obs	1,026	1,026	1,026	1,026	1,026
R^2	0.7800	0.7568	0.8272	0.8429	0.9317
Mean	1.7380	1.0845	0.3978	0.2557	0.7443
Median	1.5170	0.9248	0.3412	0.2133	0.9763
Std. Dev.	1.0615	0.6872	0.2714	0.1955	0.3278

The table reports the results of regressing measures of public goods on the direct British rule dummy, per the following specification.

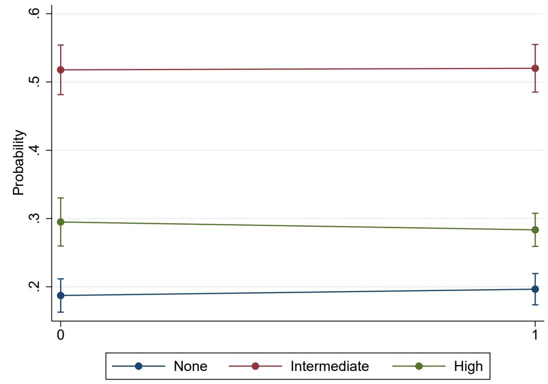
$$Y_{j,t} = \beta \cdot \text{Direct Rule}_j + \theta_{p(j\in p)t} + \delta_1 \cdot \text{Latitude}_j + \delta_2 \cdot \text{Longitude}_j + \epsilon_{j,t}$$

where $Y_{j,t}$ denotes the outcome variables observed for a district j at time t. $Y_{j,t}$ is the number of schools per 1,000 people in column 1, number of primary schools per 1,000 people in column 2, number of middle schools per 1,000 people in column 3, number of secondary and senior secondary schools per 1,000 people in column 4, and the fraction of villages with electricity connection in column 5. Direct Rule (=1) is a binary variable that takes a value of 1 if the district was under direct British rule before independence in 1947, and 0 otherwise. $\theta_{p(j\in p)t}$ denotes the district-pair × year fixed effect. Lat/Long indicates controls for latitude and longitude of the district. The unit of observation is district year and the data on schools and electricity connection comes from the Census data for 1991, 2001 and 2011. Standard errors reported in parentheses are two-way clustered by native states and district-pair. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

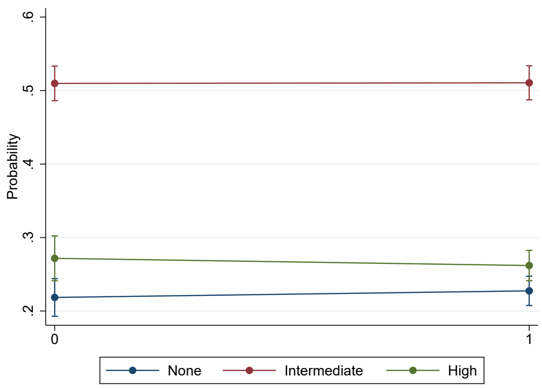
Figure F.2: Trust in State Apparatus



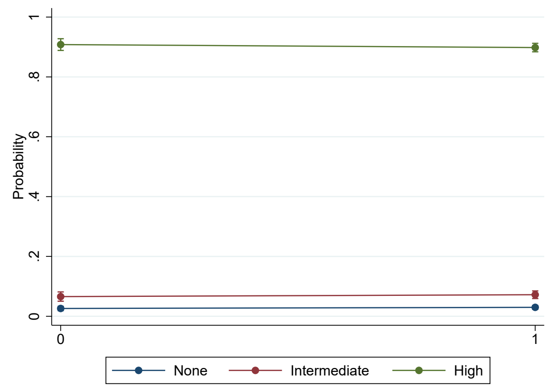
(a) Trust in Politicians



(b) Trust in Panchayats



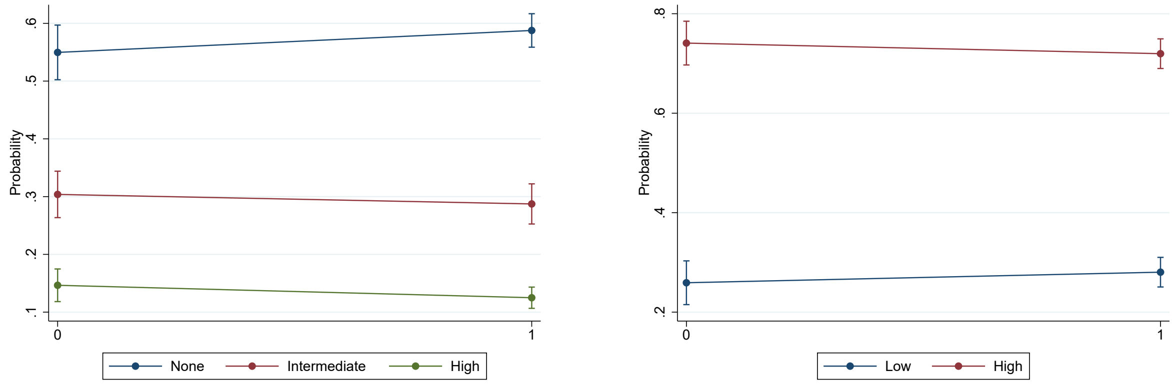
(c) Trust in Police



(d) Trust in Banks

The figure plots the average probability of response for trust in different segments of the state apparatus. The average probabilities are computed using an ordered probit regression. We control for the gender, age, religion/caste of the respondent and account for the location being urban or metro. In addition, we include state fixed effects. 0 denotes indirect rule and 1 denotes direct rule. The respondents are asked - "How much confidence do you have in X?" X denotes politician, panchayats, police and banks. A respondent can either say - (1) *No Confidence* coded as "None", (2) *Intermediate level of confidence* coded as "Confidence," or (3) *High level of confidence* coded as "High." The sum of the average three probabilities within a location add up to 1.

Figure F.3: Community Cooperation & Conflict



(a) Community Conflict

(b) Community Cooperation

The figure plots the average probability of response for general conflict and community cooperation. The average probabilities are computed using an ordered probit regression. 1 denotes direct rule and 0 denotes indirect rule. We control for the gender, age, religion and caste of the household head and account for the location being urban or metro. In addition, we include state fixed effects. The respondents are asked - "How much general conflict is their in your neighbourhood or village?" for panel F.3a. A respondent can either say - (1) *No Confidence* coded as "None", (2) *Intermediate level of confidence* coded as "Intermediate," or (3) *High level of confidence* coded as "High." The sum of the average three probabilities within a location add up to 1. The respondents are asked - "How much general cooperation is their in your neighbourhood or village?" for panel F.3b. A respondent can either say - (1) *High* coded as "High" or (3) *Low* coded as "Low." The sum of the average the two probabilities within a location add up to 1.