What Explains Geographic Variation in Corporate Investment?*

Shohini Kundu† Nishant Vats‡

May 13, 2021

Abstract

We show that institutions can explain geographic concentration over and above agglomerative forces. This paper uses within country variation in institutional quality, combined with a local identification approach and instrumental variable strategy to explain spatial differences in investment. We use direct (indirect) British rule as proxies for areas with low (high) institutional quality. Institutions can explain 13% of total geographic variation in investment. Moreover, investment is 8-10% lower in areas with low institutional quality. Differences in institutional quality manifest as greater court delays, impeding contractual claims, property rights, and dispute resolution, thereby decreasing investment and increasing project abandonment.

*We thank Rainald Borck, Francesco D'Acunto, Zhiguo He, Lakshmi Iyer, Ralph Koijen, Yueran Ma, Nagpurnanand Prabhala, Raghuram Rajan, James Robinson, Amir Sufi, Michael Weber, and Eric Zwick for helpful comments and suggestions. We are thankful to the Fama-Miller Center at the University of Chicago, the NSE-NYU Stern Initiative on the Study of Indian Capital Markets, and the Wheeler Institute for Business and Development at the London Business School for financial support. We thank seminar participants at the 10th Urban Economic Association Meeting and the NSE-NYU Conference on Indian Financial Markets, 2020. We do not have any conflicts of interest to disclose. We take responsibility for all errors.

†Shohini Kundu is at the Booth School of Business, University of Chicago. e-Mail: skundu@chicagobooth.edu

‡Nishant Vats (corresponding author) is at the Booth School of Business, University of Chicago. Send correspondence to 5807 S Woodlwan Ave, Chicago, IL, USA. e-Mail: nvats@chicagobooth.edu

1 Introduction

What explains geographic concentration in investment? Do institutions play a role? 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: institutions. We consider how the circumstances in which agglomerative forces take hold can influence economic outcomes. Specifically, we emphasize the importance of *rules of the game* (institutions), the context in which factors of production operate.

The relationship between institutional quality and investment remains unclear due to a host of endogeneity issues. For example, better institutions can facilitate investment by reducing the holdup problem. However, greater economic activity and human capital accumulation – both correlated with greater investment – may also foster better institutions resulting in a simultaneity problem. The extant empirical literature presents mixed evidence supporting both hypotheses.[1] These studies draw inference based on cross-country analyses which are susceptible to several endogeneity issues such as omitted variable bias, selection bias, simultaneity problem, etc.

In this paper, we attempt to address these endogeneity issues and show that institutions can explain geographic concentration over and above the traditional agglomerative forces. We do this by using a two-pronged approach. First, we combine within country analysis

[1]The causal relation between institutions and economic growth is ambiguous. On the one hand, Hall and Jones (1999), Acemoglu, Johnson and Robinson (2001), Acemoglu, Johnson and Robinson (2002), Easterly and Levine (2003), Dollar and Kraay (2003), and Rodrik, Subramanian and Trebbi (2004), among others show that institutions cause economic growth. On the other hand, Barro (1999), Przeworski et al. (2000), Djankov et al. (2003), and Glaeser et al. (2004), among others argue that growth in income and human capital causes institutional improvements.

with a plausibly exogenous measure of institutional quality based on the historical origins of an area. 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 omitted variables bias, 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 institutions and economic activity, in general, and, in particular, investment concentration.

This paper uses the within-country geographic variation in institutional quality 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, parts of the Indian subcontinent fell under the "direct rule" of the British or "indirect rule," under the administration of Indian rulers. While the British directly controlled the direct ruled areas in India, Indian rulers had considerable autonomy over the internal affairs of their kingdom in indirect ruled areas. Iyer (2010) argues that the kings in indirect ruled areas were under a constant threat of deposition by the British on account of misrule. Hence, local kings exerted tremendous effort to improve institutional quality, to avoid the slightest hint of misrule. While the precolonial geographic differences in de jure institutions were eliminated after the Indian Independence of 1947, we argue and verify that precolonial geographic differences in de jure institutional quality affect contemporary de facto institutional quality, especially contract enforcement, property rights protection, and dispute resolution. Therefore, India provides an ideal laboratory to examine the consequences of differences in institutional quality on geographic variation in investment concentration.

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 within-state investment concentration is 20 percentage points higher relative to any of the benchmarks, and the proportion of districts under direct rule can explain 13% of total variation in within-state investment concentration.

States with a larger 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 × year and state × 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 suggests that the baseline results are unlikely to be driven by omitted variables.

A direct comparison of direct and indirect ruled districts may bias our inference in the presence of systematic differences between direct and indirect ruled districts. Our baseline analysis assumes that one district in India is as good a control as any other district within a state. We validate this assumption by examining differences across several observable characteristics, such as geography, distance from state capital, the religion of precolonial rulers, and the colonial population composition.

In spite of these tests, we cannot completely rule out differences among districts along all observable and unobservable dimensions. 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 × district-pair × 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, and investment opportunities 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 two instrumental variable strategies.

First, we exploit a unique feature of British annexation policy in India, the Doctrine of Lapse. The Doctrine of Lapse was in effect between 1848 and 1856, which allowed the governor-general of British India to annex Indian princely states where the ruler died without a natural male heir. The relevance condition posits that the death of a ruler without a natural male heir is associated with a territory coming under direct British rule. The exclusion restriction posits that the death of a ruler without a natural male heir affects current corporate investment only through the territory being under direct British rule. We verify the relevance and exclusion restrictions associated with this instrument in the first stage, and a falsification exercise using the death of the ruler without a natural male heir in periods when the policy was not applicable. 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.

However, the instrument of Doctrine of Lapse only allows us to compare direct ruled districts annexed during a specific period with all indirect ruled districts. This raises concerns about the validity of the estimate obtained using direct ruled districts acquired during a specific period to all direct ruled districts in India. We address this concern using a second instrument that is immune to such a critique. We use a district's cotton cultivation during precolonial times as an instrument for direct British rule. We validate the relevance assumption in the first stage of our estimation. The significant first stage result suggests that the British sought control of cotton-producing districts. This instrument not only overcomes the "temporal" policy critique of using Doctrine of Lapse as an instrument, but has sufficient variation allowing us to include firm $\times$ district-pair $\times$ year fixed effect, comparing district-pairs that are likely to be similar across several observable and unobservable dimensions. The 2SLS estimation applies a local identification approach, using precolonial cotton cultivation in a district as an instrument for direct rule. The estimates from this approach are similar in magnitude to the estimate produced using the death of ruler without a male heir as an instrument for direct British rule. The two instrumental variable strategies confirm the magnitude of the estimate, allowing us to interpret it as causal.

We explore the underlying mechanism through which direct British rule affects investment. The threat of British annexation under the pretext of misrule pressured Indian rulers to create better institutions. The long-tenure of Indian rulers, combined with the motivation to create a legacy, further incentivized them to invest for the long-term. This led to the emergence of colonial differences in institutional quality between direct and indirect ruled districts. Though all districts within a state, regardless of their colonial history, were brought under the umbrella of an identical legal and administrative framework at independence, the colonial differences across districts persists and affects current institutional quality (Iyer (2010)). Hence. we argue that the long-term differences between direct and indirect ruled districts are driven by differences in institutional quality. Further, we rule out two alternative channels that may explain differences in investment, namely, differences in mistrust of the state, and differences in community conflict.

We focus on a feature of institutional quality fundamental to investment: contract enforcement. To this end, we study the differences in case delays across direct and indirect ruled districts. Fast and effective processing of cases is important for reducing uncertainty and ensuring that individual and stakeholder rights are well protected. Our local identification approach comparing similar legal cases across district-pairs indicates that courts in direct ruled districts take 10% longer, on average, to resolve similar cases relative to courts in indirect ruled districts. Case delay can impede protection and enforcement of property rights, thereby, exposing firms to greater uncertainty. Greater uncertainty reduces investment. Additionally, we show that it increases the incidence of project abandonment.

Lastly, we provide external validity for our mechanism. We start by documenting that global foreign investment is concentrated across countries. We employ several measures of institutional quality, including settler mortality of Acemoglu, Johnson and Robinson (2001), and other measures of contract enforcement, property rights protection and swift dispute resolution such as government effectiveness, regulatory quality, rule of law, and control of corruption. The evidence consistently shows that institutional quality can explain the geographic concentration of investment across countries.

Related Literature: This paper is primarily related to three strands of the literature: geographic concentration of economic activity, the relationship between institutional quality and investment, and 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: institutions. 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)). In this paper, we argue that institutions can explain geographic concentration of investment over and above these traditional agglomerative forces. Our local identification approach using firm × district-pair × year fixed effects allows us to disentangle the importance of differences in institutional quality from differences in traditional agglomerative forces of Marshall (1920), differences in geographic features, and differences in investment opportunities.

Second, we contribute to the academic debate on the relationship between institutions and economic growth by addressing endogeneity concerns. The extant literature studying the causal relation between institutions and economic growth, using cross-country analyses have documented mixed results. On the one hand, Hall and Jones (1999), Acemoglu, Johnson and Robinson (2001), Acemoglu, Johnson and Robinson (2002), Easterly and Levine (2003), Dollar and Kraay (2003), and Rodrik, Subramanian and Trebbi (2004), among others show that institutions cause economic growth. On the other hand, Barro (1999), Przeworski et al. (2000), Djankov et al. (2003), and Glaeser et al. (2004), among others argue that growth in income and human capital cause institutional improvements. In contrast to previous research, we conduct a within-country analysis. We address issues of omitted variable bias, selection bias, and simultaneity bias by employing a local identification approach, using firm × district-pair × year fixed effects, and two instrumental variable strategies. Hence, our empirical strategy is better adept at solving the difficult identification problems at the heart of the debate.

Third, we are also related to the longstanding literature, exploring the long-run effects of colonial rule by identifying the mechanism through which colonial regimes can affect contemporary economic outcomes. Prior research has examined several mechanism through with colonial regimes persist such as mistrust (Nunn and Wantchekon (2011)), economic and physical infrastructure (Donaldson (2018), Dell and Olken (2020), Berger (2009)), and health and education infrastructure (Huillery (2009)), among others. Our work is closest to the set of papers that identify property rights as the primary mechanism of long-run persistence of colonial rule. However, we deviate from previous work which argues colonial institutions influenced current legal regimes for property protection, e.g., Acemoglu, Johnson and Robinson (2001), Porta et al. (1998), Banerjee and Iyer (2005). We attribute the differences in institutional quality to differences in the enforcement of the law, as measured by court case delays – not differences in the law itself. Our within state estimation strategy explicitly controls for all differences in the law. Therefore, our work provides another mechanism through which colonial institutions influence institutional quality.

Our work builds on previous work examining the long-run consequences of colonial rule, in particular, Iyer (2010). However, our work fills an important gap in the literature, as we identify potential sources of differences in institutional quality, not studied in Iyer (2010). 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. However, 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 and unlikely to explain the persistent differences in investment concentration even more than two decades after 1991. We rather argue that differences in investment concentration are likely to be driven by differences in court delays across direct and indirect ruled district.

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 robustness results. Section 6 presents the evidence from bordering districts. Section 7 presents the results from the instrumental variable regressions. Section 8 discusses the mechanism through which institutional quality affects investment. Section 9 presents the results, demonstrating external validity, and section 10 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. Iyer (2010) argues that the districts that fell under indirect rule, where precolonial rulers retained administrative powers, had higher institutional quality, while districts that fell under direct rule of the British had lower institutional quality.

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, Subsidiary Alliance, the Doctrine of Lapse, 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 male 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.[2] 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

[2]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.

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. 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.

2.2 Why do Colonial Differences Persist?

The long-term effects of differences between direct and indirect British rule originate from several sources. First, 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 institutions. Secondly, 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. Thus, the combination of a longer tenure, as well as the threat of cession contributed to development of better early institutions and governance policy in indirect ruled areas, resulting in higher institutional quality. These colonial differences in institutional quality are likely to persist. On the one hand, wealth-maximizing postcolonial elites inheriting extractive institutions may not invest in improving these institutions. On the other hand, if the costs of creating better institutions have been sunk by precolonial elites, it may not pay off for the new postcolonial elites to demote them (Acemoglu and Verdier (1998); Acemoglu, Johnson and Robinson (2001); Acemoglu and Robinson (2005)).

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 4 presents results from a linear probability model, regressing the likelihood of an area being under direct rule on these characteristics. Table 4 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 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 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.[3] 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.[4] 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 dispersion within 427 districts in the sample. Appendix B.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

[3]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.

[4]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.

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.

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

3.3 Judiciary Data

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

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

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 25th and 75th percentiles of case duration are 28 days and 746 days, respectively. Moreover, the total number of cases also exhibit a large heterogeneity.

3.4 Global Data on Investment and Institutional Quality

We employ several datasets on worldwide investment and institutional quality. We use the IMF's Coordinated Direct Investment Survey (CDIS) data for computing a country's share of total direct investment. CDIS provides detailed data on "inward" and "outward" direct investment, i.e., direct investment into and out of the reporting economy, respectively for all reporting economies. This data is available from 2009-2019. We focus our analysis on outward investment by G20 countries, which includes Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, South Korea, Mexico, Russia, South Africa, Turkey, United Kingdom, and the United States.[6] We use data on settler mortality as a proxy of institutional quality from Acemoglu, Johnson and Robinson (2001). Supplementary time-varying measures of institutional quality come from the World Bank Governance Indicators (WGI). WGI reports governance indicators for all reporting economics between 1996-2019. In particular, we focus on four indicators: government effectiveness, regulatory quality, rule of law, and control of corruption.[7]

4 Results

This section describes the key results, and the associated empirical methodology. 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.

[6] There is no data available for Saudi Arabia.

[7] We direct the readers to Kaufmann, Kraay and Mastruzzi (2011) for a discussion on the methodology of the Worldwide Governance Indicators (WGI) project, and related analytical issues.

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. This allows us to examine the relationship between the concentration of corporate investment and de facto institutional quality, while holding the law fixed.

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 in 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 likely to be 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 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 displays the baseline specification:

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct rule}_j + \theta_{i,y} + \theta_{s(j \in s),y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t} \tag{1}$$

$Ln(Y_{i,j,t})$ 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.[8] 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. The β estimate may be interpreted as causal under the assumption of randomness of Direct Rule$_j$. We discuss this assumption in detail in section 7.

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

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

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 district is ₹ 91.19 million. This amounts to an average within-state difference of ₹ 8.81 mmillion per project across districts.[9] 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 detail next.

Moreover, we validate the precision of the estimates through the Conley (1999) standard errors, adjusting for spatial dependence within 100 km, reported in square brackets. Overall, the results indicate the estimate is negative and statistically significant. 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

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

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 $\times$ time fixed effects 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 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). 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.

5 Robustness

5.1 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 sharply 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 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. Thus, we reject the hypothesis that the relationship between project size and historical origin of a district is driven by omitted variables.

5.2 Controlling for Other Geographic Covariates

This section examines the robustness of the baseline estimate to other time-invariant district level covariates. The British were likely to take over strategic territories. This includes territories along the coast, territories with greater geographic area, or territories that have favorable climatic conditions. To control for this, we include geographic covariates, which may be correlated with the likelihood of a district being under direct rule. In addition to geographic covariates, we control for the distance of the district from the state capital, as a state's capacity to provide necessary infrastructure for investment may be restricted by the physical distance between administrators and citizens (Bardhan (2002); Asher, Nagpal and Novosad (2018)).

Table 5 reports the estimation results of the baseline equation 1 after controlling for geographic covariates. The estimate of interest is the coefficient of Direct Rule$_j$, β. The estimate of β is negative, statistically significant at the 5% level and remains relatively stable in magnitude across columns (1)-(9). The magnitude of β in column (9) is quantitatively similar to the magnitude of β in column (5) of table 3. The vast majority of geographic covariates cannot explain the within state heterogeneity in corporate investment. However, altitude, rainfall and distance to state capital are statistically significant. Higher altitude, rainfall, and greater distance to state capital are associated with lower investment. These results are consistent with the notion that unfavorable geographic conditions and distance from state headquarters may deter corporate investment. Overall, the results indicate that the estimate of β is not driven by other geographic characteristics that could potentially be correlated with the direct rule variable.

5.3 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. 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.[10]

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

[10]Section C discusses the calculation and the magnitude of spatial autocorrelation of direct British rule.

this as placebo direct rule. We estimate the coefficient associated with the placebo direct rule and repeat this exercise 10,000 times. To negate the validity of 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 in investment.

6 Evidence from Bordering Districts

The traditional approach used to compare the differences in investment level 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 in cases where one district was 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 districts are immediately adjacent and are expected to be similar in both observable and unobservables characteristics and are likely to follow similar paths had India not been colonized. Moreover, whether a district within a contiguous direct-indirect ruled pair was 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)) as these parameters are likely to be similar across contiguous districts.

6.1 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:

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,p(j \in p),y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t} \qquad (2)$$

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.

6.2 Results

Table 6 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 accounts for all 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 similar size. Figure 6b presents the sample. The indirect ruled districts, referred 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 7 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 6 are unlikely to be driven by systematic differences across contiguous districts.

7 Instrumental Variable Regression

This section reports and discusses the results from two instrumental variable (IV) strategies to address the selection issue. Our first strategy uses the death of the ruler without a male heir, and the second strategy exploits the fact whether an area was a cotton producing area during the precolonial period, as an instrument for direct British rule. Our two-stage least square estimates validate our baseline results, ruling out the concern that our results may be plagued by selection bias.

7.1 Instrument: Death of Ruler, without a Male Heir

This section reports and discusses the results from an instrumental variable (IV) regression where we instrument for direct British rule, using the British policy of the Doctrine of Lapse. Under the Doctrine of Lapse, the governor-general of British India annexed the Indian princely states where the ruler died without a natural male 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 male heir, during 1848 to 1856, as an instrument for a territory coming under the direct British rule.

Thus far, we compare the 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 which hinders interpreting the estimate of β in equation 1 as causal. 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, under the Doctrine of Lapse. The instrument relies on two crucial assumptions: relevance and exclusion. First, the relevance restriction posits that the death of a ruler without a natural male heir can explain the likelihood of a territory coming under the direct British rule. Second, the exclusion restriction posits that the death of a ruler without a natural male heir affects current corporate investment only through the territory being under direct British rule. The first assumption is a mechanical byproduct of the policy, which we verify in the first stage.[11] The identifying assumption required for the exclusion restriction to hold is that the death of the ruler without a natural male 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 identifying assumption in section 7.1.1.

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 1847 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. The magnitude of the estimate is, however, greater for this sample. Column (2) presents the IV specification 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, respectively. Together, these form a 2SLS estimation strategy. The instrument is positively related to the likelihood that a district is under direct British rule. A district whose ruler died without a natural male heir between 1848 and 1856 is 66% more likely to be under direct British rule. The f-statistic of the first stage is 28.80 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 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 are randomly distributed in space. Overall, the results indicate that the results discussed so far are unlikely

[11] 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 male 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.

to be driven by selection.

7.1.1 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 male 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 male heir between 1848 and 1856 is likely a matter of chance, and unlikely to have a direct impact on the corporate investment in the present.

The identification strategy is invalid if the Doctrine of Lapse was strategically implemented to acquire certain areas, or rulers could choose to opt out of this policy. The Doctrine of Lapse did not allow the British to take over their most desired states of Oudh and Hyderabad. Further, 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 out of the policy (Iyer (2010)). Additionally, the administration of all states under direct British rule was identical irrespective of the mode of annexation.

Exclusion requires that the death of a ruler without a natural male heir affect current investment by only influencing the likelihood of direct rule during colonial times. In this paragraph, we consider several threats to the validity of the exclusion restriction. The death of a ruler without a natural male heir could be attributable to several other factors, e.g., environment, genetics, etc., which may effect postcolonial corporate investment through other channels. Alternatively, the death of a ruler without a heir may can throw a territory into a period of political uncertainty, resulting in long-run consequences captured by our estimation. We address these two claims via a falsification test. We use the death of a ruler without a natural male heir in the period of 1858 and 1884 as a false instrument.

The policy of Doctrine of Lapse was officially abandoned after 1858. Therefore, the false instrument should show a negligible effect under the exclusion restriction assumption. Table 9 reports these results. The IV estimate reported in column (1) is statistically insignificant and very small in magnitude relative to the IV estimate reported in column (2) of table 8. The first stage estimation using the false instrument results in a small, closer to 0, and statistically insignificant first stage coefficient. The f-statistic associated with the first state is 1.11 implying that the false instrument is a weak instrument. The second stage estimate of direct rule using the false instrument is statistically insignificant and has a positive sign, contradicting the predictions of the two threats posed to the exclusion restriction.[12]

7.2 Instrument: Cotton Producing Territories

In the previous section, we use the death of a ruler without a male heir as an instrument for direct British rule, allowing us to infer the differences in direct and indirect ruled districts based on a subsample of direct ruled districts that were annexed under the direct British rule from 1848 to 1856. Inference based on a subsample may still be plagued by selection issues if the British annexation policy and subsequent colonial institutions changed with the timing of annexation. We address this concern by using an alternate instrument that is immune to such a critique.

We use a district's cotton cultivation during precolonial times as an instrument for direct British rule. The British took over direct control of cotton producing areas to control the supply of cotton, a valued commodity with a rising market share in the textile industry (Schoen (2009)). This instrument is not only immune to the temporal policy critique but has sufficient variation allowing us to include firm × district-pair × year fixed effect. Hence, this strategy allows us to address selection issues while comparing district-pairs that are likely to be similar across several observable and unobservable dimensions. Table 10 reports the results of using precolonial cotton production by a district as an instrument for direct British rule. The first stage estimate is significant and positive. The f-statistic of 47.4 in column

[12]The magnitude of the point estimate is large. This is likely to be the case given the instrument being used is a weak instrument. A weak instrument artificially inflates the point estimate of interest via a very small first stage estimate.

(6) implies relevance of the instrument. The second stage estimate reported in column (5) is negative, statistically significant, and quantitatively similar to the magnitude of the estimate reported in column (3) of table 8. Overall, our two instrumental variable strategies taken together confirm the magnitude of the estimate, allowing us to address the issue of selection bias and interpret the effect of direct British rule on investment as causal.

8 Mechanism

This section discusses the underlying mechanism through which institutional quality affects investment. Specifically, we show that courts in districts under direct British rule take longer to resolve similar cases relative to courts in districts under indirect British rule. We argue that longer judicial duration impedes contract enforcement that exposes firms to uncertainty, thereby decreasing investment. Moreover, we rule out alternative explanations relating to trust and conflict.

8.1 Contract Enforcement

Contract enforcement, property rights, and dispute resolution are essential features of well-functioning societies. High institutional quality is an important determinant of these dimensions of governance. For example, fast and effective deliverance of justice can reduce uncertainty, ensure that individual and stakeholder rights are protected, and assure the progress of society. Low institutional quality can hamper these aims. Specifically, in our context, the lack of strong institutions, can expose investors to uncertainty, thereby decreasing investment (Bloom, Bond and Van Reenen (2007)). This section documents the relationship between direct and indirect rule and a segment of institutional quality that affects contract enforcement, property rights and dispute resolution: the time taken by local courts to deliver the verdicts.

We focus on cases at local courts, as the judiciary is the main instrument for economic agents to resolve their disputes. We use data on the universe of court records to extract a measure of the time taken by courts to deliver a verdict: case delay. We aggregate the data at the district-statute-year level using the median value of case delays based on all cases filed in that year under a statute. A fast dispute resolution system has the merits of reducing uncertainty related to economic transactions, making the case duration an ideal proxy for identifying delays in conflict resolution between economic agents. Furthermore, we use the subsample of contiguous direct and indirect ruled districts shown in figure 6a to rule out issues related to omitted variable bias.

Table 11 reports the results of regressing the natural logarithm of one plus median delays at district-statute-year level on Direct Rule variable using a sample of contiguous direct and indirect ruled districts. The estimation procedure allows us to control for District-Pair $\times$ Statute $\times$ Year fixed effects. Hence, the difference in court delays between direct and indirect ruled districts is estimated by comparing a case filed under the same statute in a direct ruled district with an indirect ruled district within a contiguous district pair. The estimate of interest is positive and statistically significant. Our preferred specification, reported in column (3), indicates that courts take 10% more time to resolve cases, filed under the same statute, in direct ruled districts relative to a contiguous indirect ruled district. This implies that contract enforcement and mechanisms of dispute resolution and property rights protection are weak in direct ruled districts, increasing uncertainty around contractual and property rights.

8.2 Project Abandonment and Uncertainty

We further explore the underlying mechanism, by examining the relationship between direct rule and the status of announced projects. If indeed greater court delays in direct ruled districts impede protection and enforcement of property rights, and exposes firms to greater uncertainty, we should not only expect lower investment in these areas, but also greater project abandonment. Using our local identification approach, we find that the same firm is more likely to abandon its projects in a direct ruled district relative to a contiguous indirect ruled district. Figure 7 plots the probability of abandoning a project in a direct ruled district relative to a contiguous indirect ruled district by the same firm until five years after announcement. Our results show that the same firm is three percentage points more likely to abandon its projects in direct ruled district relative to a contiguous indirect ruled district. This is economically large, given that the unconditional probability of project abandonment is 13%. These results indicate that greater uncertainty due to long delays in enforcement not only reduces investment ex-ante, but also increases ex-post project abandonment.

8.3 Addressing Alternative Explanations: Trust & Conflict

This section discusses two alternative explanations that can potentially explain our results – trust and conflict. We argue that none of these forces can account for our results.

8.3.1 Trust

A popular narrative puts the direct exploitation of subjects at the heart of the long-run consequences of colonial rule. 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 relative 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 to invest in basic infrastructure required for corporate investment. Therefore, an alternative hypothesis is that the current difference in investment can be attributed to 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 the state actors, 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 D.1). Hence, we argue that differences in trust, due to colonial rule, is unlikely to drive our results.

8.3.2 Cooperation and Conflict

The British followed on a policy of *divide and rule* to govern India, pitting communities against each other. Hence, areas under the direct British rule could have greater degree of community conflict relative to indirect ruled districts. Therefore, an alternative hypothesis is that differences in community cooperation across direct and indirect ruled districts rather than differences in institutional quality may explain our results, as conflict increases uncertainty, thereby reducing investment. Using the 2011 survey data from the Indian Human Development Survey (IHDS) on community cooperation and conflict, we do not find any evidence of systematic differences in community cooperation or conflict across direct and indirect ruled districts (see, appendix figure D.2). Hence, we argue that differences in community cooperation or conflict, due to colonial rule, is unlikely to drive our results.

9 External Validity

In this section, we provide external validity for our mechanism. We extrapolate our within-country analysis to the world, to provide suggestive evidence that country-specific institutional quality can explain its share of global investment.

We begin our analysis by documenting geographic concentration of direct global investment. We use country-pair data on outward investment by G20 countries. The data provides information on the amount of direct investment by each G20 country into all other countries from 2009 to 2019. We compute the share of direct investment in each country, by all G20 countries relative to the total global direct investment by all G20 countries. Figure 8 presents a heat map showing the share of investment received by each country. Similar to previously documented geographic concentration of investment within India, figure 8 shows that direct investment by G20 countries also exhibits geographic concentration. Higher income countries receive a higher share of investment relative to lower income countries.

Next, we examine if the geographic concentration of the share of investment received by a country can be explained by institutional quality. To this end, we use several proxies for institutional quality. Similar to the within-country results presented thus far, the cross-country analysis shows that the share of investment received by a country increases as the institutional quality increases.

We use settler mortality as a measure of institutional quality, à la Acemoglu, Johnson and Robinson (2001). Acemoglu, Johnson and Robinson (2001) posit that settler mortality affected settlements, affecting early institutions which in turn, have persisted to the present, and continue to affect contemporary outcomes. In Figure 10, we find a negative relation between the share of investment and settler mortality, consistent with our mechanism. Next, we study whether this relationship is significant and robust to additional variables and region fixed effects in table 12. As a baseline comparison, we do not include any controls in column (1). In columns (2) to (5), we sequentially add latitude, forced labor, land area, and population controls, respectively. Columns (6) and (7) includes all controls, while column (7) additionally includes world bank region fixed effects. There are two key takeaways from this table. On the qualitative side, the estimate of settler mortality is statistically significant and negative, reflecting that institutional quality is positively associated with the share of investment. On the quantitative front, settler mortality can explain 13% of total variation in the share of investment, reflecting that the role of institutions in determining investment is non-trivial.

We extend the analysis by using alternative measures of institutional quality. Figure 10 presents binscatter plots, showing the relation between the share of investment received and four measures of institutional quality - (1) government effectiveness in figure 10a, (2) regulatory quality in figure 10b, (3) rule of law in figure 10c, and (4) control of corruption in figure 10d. Figure 10 shows that share of investment is positively correlated with these four measures of institutional quality. We test the relation between share of investment and these measures of institutional quality in a more rigorous fashion in table 13, accounting for origin $\times$ year and destination world bank region $\times$ year fixed effects. Consistent with our findings, an increase in government effectiveness, regulatory quality, rule of law, and control of corruption, shown in Columns (1) through (4), respectively, increases the share of investment.

10 Conclusion

In this paper, we show that institutions can explain geographic concentration over and above the traditional agglomerative forces. This paper uses within country variation in institutional quality combined with a local identification approach and instrumental variable strategy to explain the spatial differences in investment. We use direct and indirect British rule as proxies for areas with low and high institutional quality, respectively. Our aggregate analysis shows that institutions can explain 13% of total geographic variation in investment. Our micro-level estimate suggests that investment is 8-10% lower in areas with low institutional quality, relative to areas with high institutional quality. We further explore the mechanism through which low institutional quality affects investment. We show that differences in institutional quality manifest as greater court delays impeding individual and stakeholders' contractual claims, property rights, and dispute resolution. Overall, our results indicate that low institutional quality negatively affects investment.

Our work has three distinct contributions. First, we focus on a novel aspect that can explain the geographic concentration of investment: institutions. Second, our empirical strategy is better adept at solving the difficult identification problems associated with identifying the relationship between institutions and economic activity. Third, we present mechanism through which differences in colonial institutional quality affects economic activity.

Our findings demonstrate that institutional quality 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, and James A Robinson. 2005. *Economic origins of dictatorship and democracy.* Cambridge University Press.

Acemoglu, Daron, and Thierry Verdier. 1998. "Property rights, corruption and the allocation of talent: a general equilibrium approach." *The Economic Journal*, 108(450): 1381–1403.

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.

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

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.

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.

Asher, Sam, Karan Nagpal, and Paul Novosad. 2018. "The Cost of Distance: Geography and Governance in Rural India." *World Bank 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).

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.

Bardhan, Pranab. 2002. "Decentralization of governance and development." *Journal of Economic Perspectives*, 16(4): 185–205.

Barro, Robert J. 1999. "Determinants of democracy." *Journal of Political Economy*, 107(S6): S158–S183.

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

Bloom, Nick, Stephen Bond, and John Van Reenen. 2007. "Uncertainty and investment dynamics." *The Review of Economic Studies*, 74(2): 391–415.

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

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

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.

Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2003. "The new comparative economics." *Journal of Comparative Economics*, 31(4): 595–619.

Dollar, David, and Aart Kraay. 2003. "Institutions, trade, and growth." *Journal of Monetary Economics*, 50(1): 133–162.

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.

Easterly, William, and Ross Levine. 2003. "Tropics, germs, and crops: how endowments influence economic development." *Journal of Monetary Economics*, 50(1): 3–39.

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.

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2004. "Do institutions cause growth?" *Journal of Economic Growth*, 9(3): 271–303.

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.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2011. "The Worldwide Governance Indicators: Methodology and Analytical Issues." *Hague Journal on the Rule of Law*, 3(2): 220–246.

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

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

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

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

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.

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.

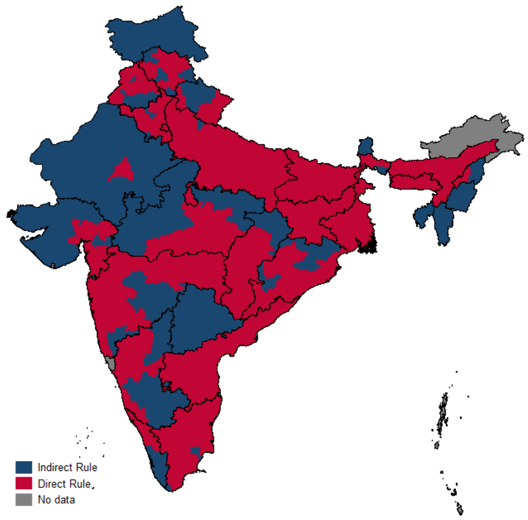
Porta, Rafael La, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert W Vishny. 1998. “Law and finance.” *Journal of Political Economy*, 106(6): 1113–1155.

Przeworski, Adam, Michael E Alvarez, Jose Antonio Cheibub, Fernando Limongi, et al. 2000. “Democracy and Development.” *Cambridge Books*.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. “Institutions rule: the primacy of institutions over geography and integration in economic development.” *Journal of Economic Growth*, 9(2): 131–165.

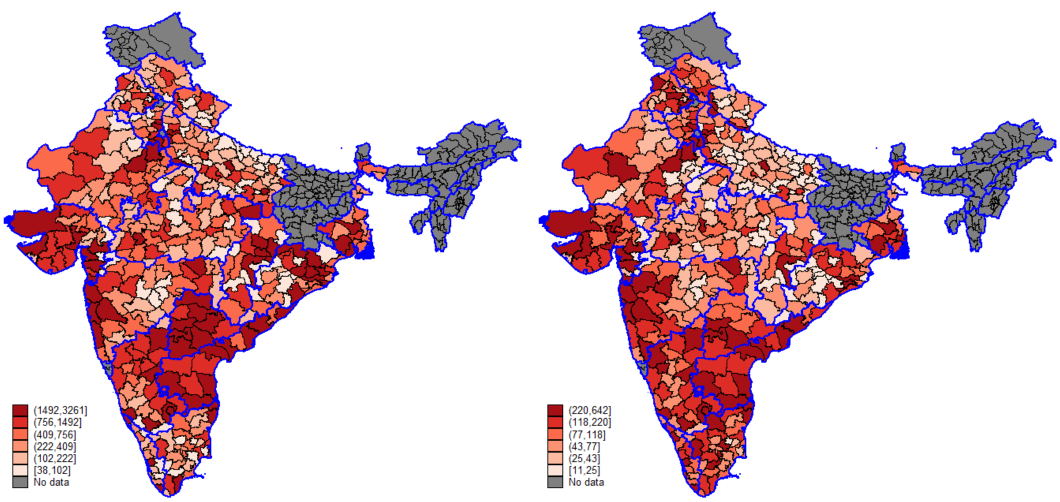
Schoen, Brian. 2009. *The Fragile Fabric of Union: Cotton, Federal Politics, and the Global Origins of the Civil War.* JHU Press.

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

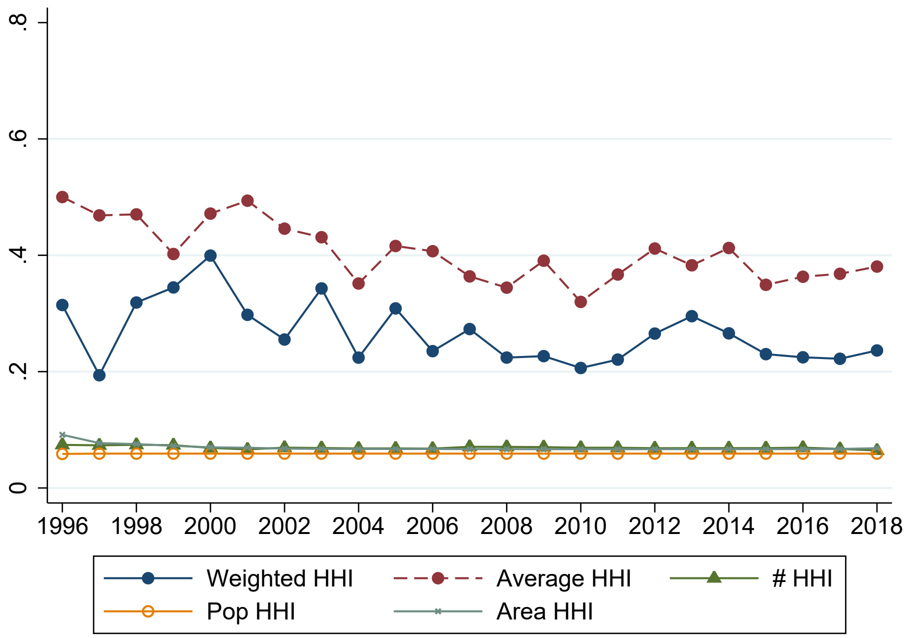


(a) Total Amount (in ₹ bn)

(b) # Projects

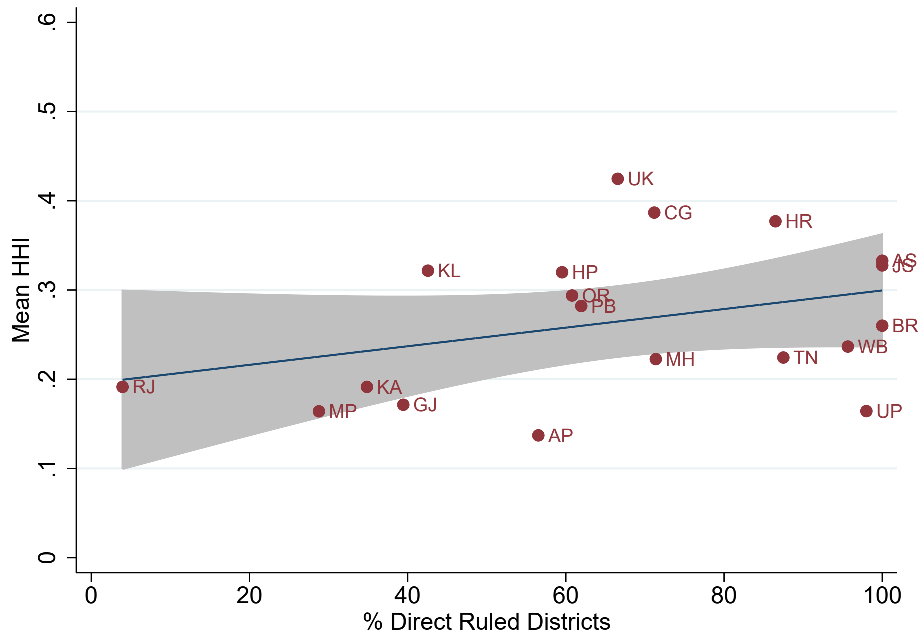
This figure plots the heatmap of total amount (in ₹ billions) and the number of projects announced in the 19 states in our sample between fiscal year 1996 and 2018. Total amount and number of projects are calculated as the total amount and total number of projects announced in each district. The exact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

Figure 3: Geographic Concentration of Investment



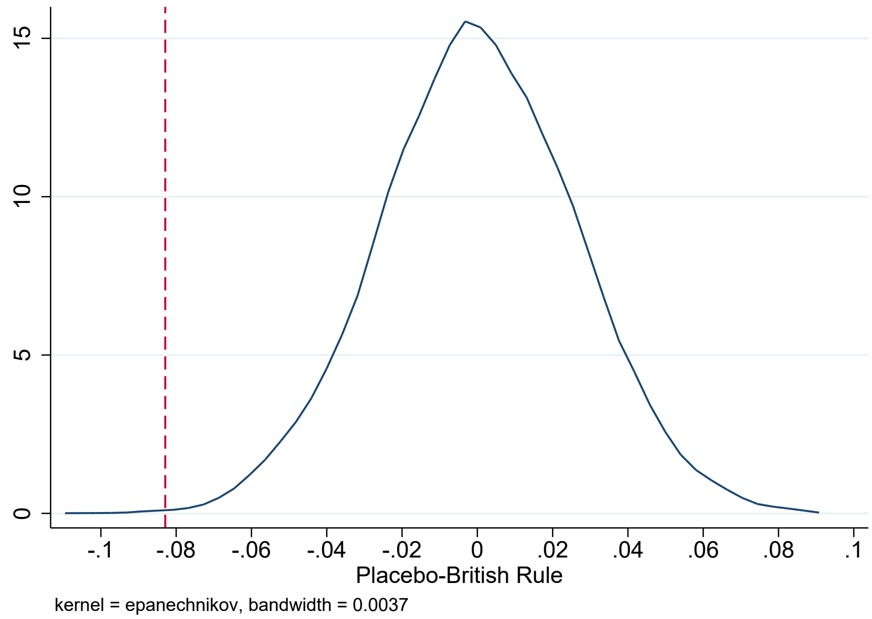
The figure plots a time-varying measure of state-level average and weighted Herfindahl-Hirschman Index (HHI). HHI is computed adding the squared share of investment in a district relative to the state in a year. The red line plots the simple average HHI for a year across all states. The blue line plots 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.

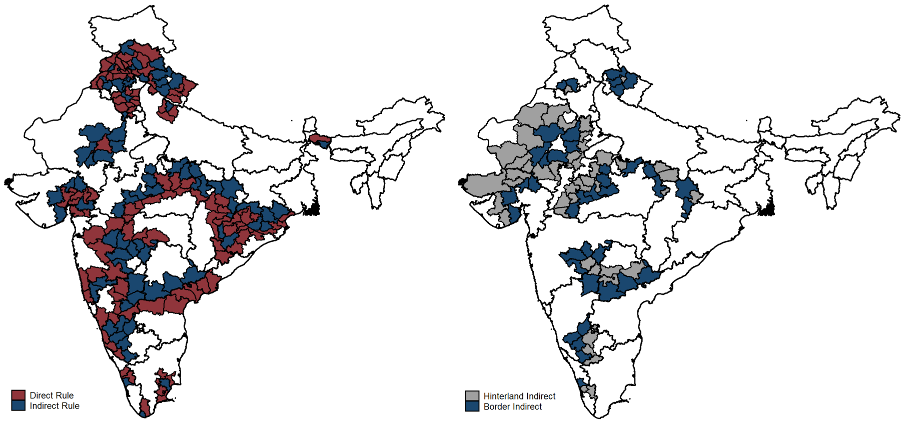
Figure 5: Placebo Test: Randomization of the Direct British Rule



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.

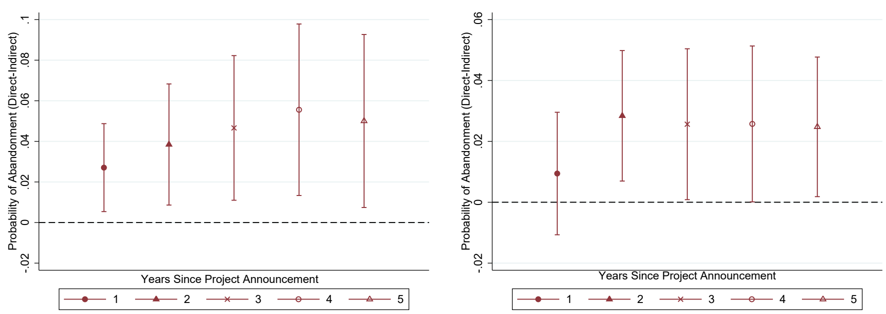
Figure 6: Sample for Local Identification Approach



(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.

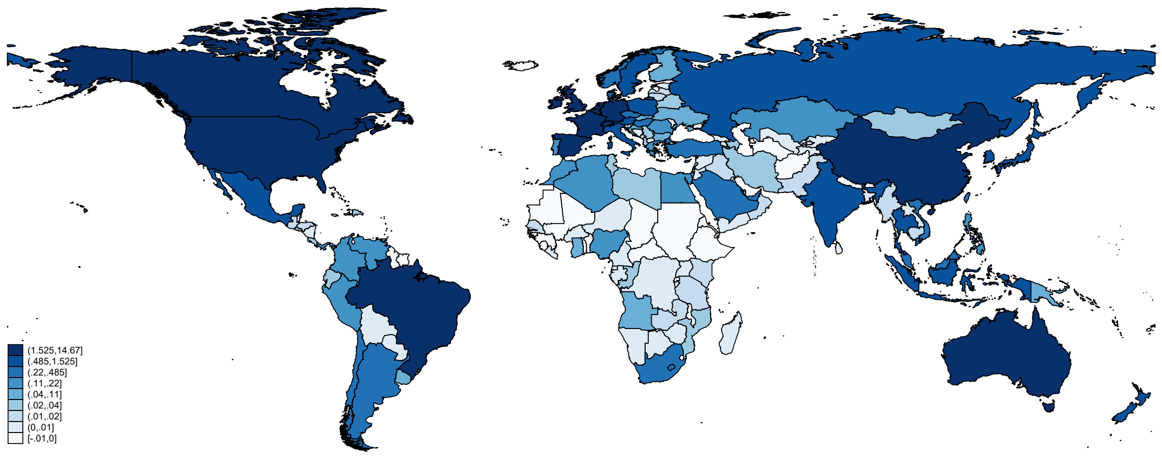
Figure 7: Project Abandonment and Direct British Rule



(a) Unconditional Estimates (b) Estimates from Local Identification Approach

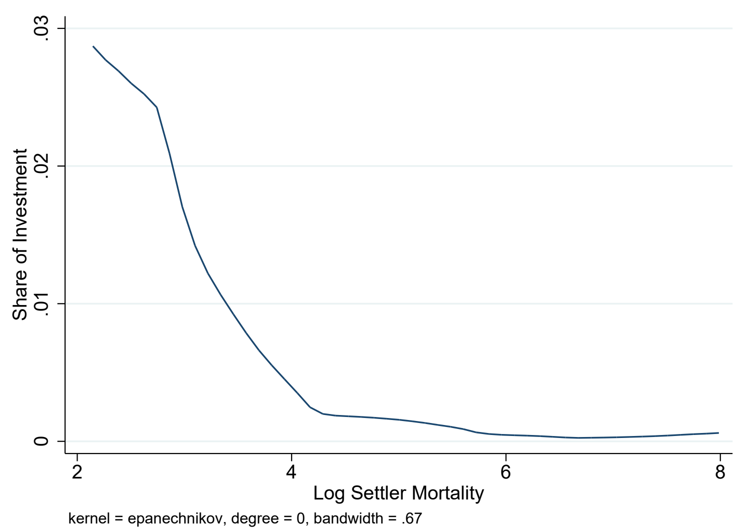
The figure plots the probability of abandonment of project announced in direct ruled districts relative to the projects announced in indirect ruled districts. We identify the status of the project each year, until five years of its announcement. Figure 7a plots the unconditional probability of abandonment of project announced in direct ruled districts relative to the projects announced in indirect ruled districts. Figure 7b plots the probability of abandonment of project announced in direct ruled districts relative to the projects announced in indirect ruled districts from local identification approach that includes firm × district-pair × year fixed effects. The 90% confidence intervals are estimated by multiway clustering the standard errors by native states, district-pair, and time of project announcement

Figure 8: Global Investment Concentration



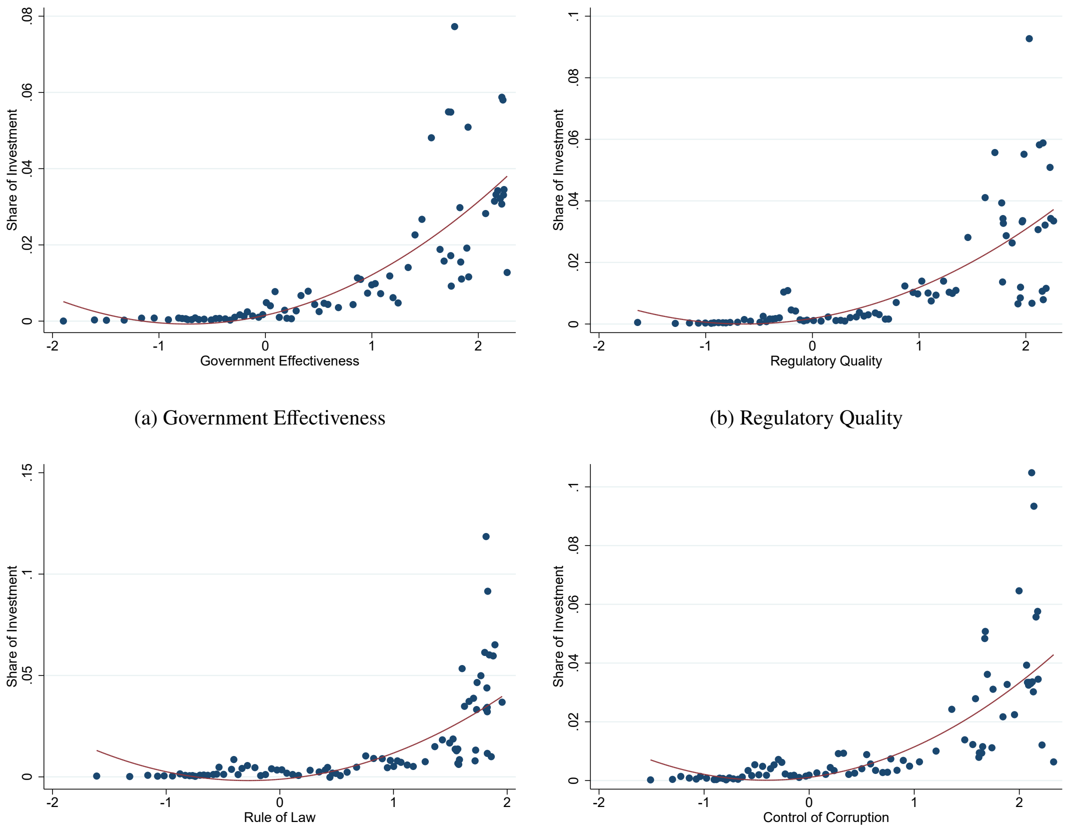
The figure presents the world map of investment share. Investment share is defined as the share of direct investment received in a country relative to worldwide direct investment by G20 countries between 2009 and 2019. The darker shading indicates greater investment share. The lighter shading indicates lower investment share. The exact international geographic boundaries have not been verified.

Figure 9: Share of Investment and Settler Mortality



The figure plots the smoothed values of a kernel-weighted local polynomial regression of investment share and settler mortality. Investment share is defined as the share of direct investment received in a country relative to worldwide direct investment by G20 countries between 2009 and 2019. The measure of settler mortality comes from Acemoglu, Johnson and Robinson (2001).

Figure 10: Share of Investment and Governance Indicators



(a) Government Effectiveness

(b) Regulatory Quality

(c) Rule of Law

(d) Control of Corruption

The figure presents the binscatter plots of investment share and various measures of institutional quality. Investment share is defined as the share of direct investment received in a country relative to worldwide direct investment by G20 countries between 2009 and 2019. The four measures of institutional quality – government effectiveness (figure 10a), regulatory quality (figure 10b), rule of law (figure 10c), and control of corruption (figure 10d) comes from the World Bank Governance Indicators.

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 the 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:

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,y} + \theta_{s(j\in s),y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t}$$

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	$I_{j,t}$	$I_{j,t} \vert I_{j,t} > 0$	$\text{Projects}_{j,t}$	$\text{Projects}_{j,t} \vert \# > 0$	$\frac{I_{j,t}}{\sum_{j \in s} I_{j,t}}$	$\frac{\text{Projects}_{j,t}}{\sum_{j \in s} \text{Projects}_{j,t}}$
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: 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:

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \Theta X_j + \theta_{i,y} + \theta_{s(j \in s),y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t}$$

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 6: 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 $\times$ Year FE	Yes	Yes	Yes
Firm $\times$ Year FE	Yes	Yes	
District-Pair $\times$ Year FE	Yes	Yes	
Firm $\times$ District-Pair FE		Yes	
Firm $\times$ District-Pair $\times$ 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:

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,p(j\in p)y} + \theta_t + Latitude_j + Longitude_j \varepsilon_{i,j,t}$$

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 $\times$ district-pair $\times$ 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 7: 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:

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Hinterland}_j + \theta_{i,p(j\in p)y} + \theta_t + Latitude_j + Longitude_j \varepsilon_{i,j,t}$$

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 8: Instrumental Variable Regression: Death of Ruler without Male Heir

Dep Var: Ln(Project Size)	(1)	(2)	(3)	(4)
	OLS	IV	2SLS	
			Second Stage	First Stage
Direct Rule (=1)	-0.2236***		-0.2239**	
	(0.0604)		(0.0960)	
Ruler Death, No Heir (=1)		-0.1475*		0.6589***
		(0.0766)		(0.1225)
Qtr $\times$ Year FE	Yes	Yes	Yes	Yes
Firm $\times$ Year FE	Yes	Yes	Yes	Yes
State $\times$ Year FE	Yes	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes
# Obs	10,293	10,293	10,293	10,293
R^2	0.7630	0.7628	0.0016	0.8646
KP LM Statistic				5.9527**
KP Wald F Statistic				28.9393

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

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{s(j\in s),y} + \theta_{i,y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t}$$

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

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 male heir. The sample consists of all districts that were consolidated under direct British rule after 1847 and districts that remained under indirect rule throughout the sample. $\theta_{i,y}$ and $\theta_{s(j\in s),y}$ denote firm $\times$ year and state $\times$ 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: Falsification Test

Dep Var: Ln(Project Size)	(1)	(2)	(3)
	IV	2SLS	
		Second Stage	First Stage
Direct Rule (=1)		0.4124	
		(2.0548)	
Ruler Death, No Heir, No Lapse (=1)	-0.0207		-0.0503
	(0.0985)		(0.0557)
Qtr × Year FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes
# Obs	10,293	10,293	10,293
R^2	0.7627	-0.0114	0.7644
KP LM Statistic			1.1655
KP Wald F Statistic			0.8129

The table reports the results of regressing the natural logarithm of project cost on Direct Rule (=1) as per the following 2SLS specification using the ruler death without a natural male heir but during the period when Lapse was not valid as the instrument for Direct Rule (=1).

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{s(j\in s),y} + \theta_{i,y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t}$$

$$\begin{aligned}\text{Direct Rule}_j &= \gamma \cdot \text{Ruler Death-No Heir-No Lapse}_j + \theta_{s(j\in s),y} + \theta_{i,y} + \theta_t \\ &+ Latitude_j + Longitude_j + \epsilon_{i,j,t}\end{aligned}$$

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, No Lapse (=1) takes a value of 1 if the incumbent Indian ruler dies without a natural male heir but during the period of 1858 to 1884 when the Doctrine of lapse was not valid. The sample consists of all districts that were consolidated under direct British rule after 1847 and districts that remained under indirect rule throughout the sample. $\theta_{i,y}$ 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 10: Instrumental Variable Regression: Precolonial Cotton Production

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	2SLS		IV	2SLS	
		Second Stage	First Stage		Second Stage	First Stage
Direct Rule (=1)		-0.3119***			-0.2272*	
		(0.1078)			(0.1270)	
Precolonial Cotton (=1)	-0.1293***		0.4144***	-0.2073*		0.9120***
	(0.0372)		(0.1119)	(0.1166)		(0.1326)
Qtr × Year Fe	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes			
State × Year FE	Yes	Yes	Yes			
Firm × District-Pair × Year FE				Yes	Yes	Yes
Lat/Long	Yes	Yes	Yes	Yes	Yes	Yes
# Obs	19,800	19,800	19,800	9,491	9,491	9,491
R^2	0.7305	-0.0026	0.6738	0.7901	-0.0009	0.7786
KP LM Statistic			10.0881***			7.3567***
KP Wald F Statistic			13.7112			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

$$Ln(Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_j + \theta_{i,p(j\in p),y} + \theta_{i,y} + \theta_t + Latitude_j + Longitude_j + \varepsilon_{i,j,t}$$

$$\text{Direct Rule}_j = \gamma \cdot \text{Cotton}_j + \theta_{i,p(j\in p),y} + \theta_{i,y} + \theta_t + Latitude_j + Longitude_j + \epsilon_{i,j,t}$$

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. 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 11: Delay and Direct British Rule

Dep Var: Ln(1+Median Delay)	(1)	(2)	(3)
Direct Rule (=1)	0.0871*	0.1048**	0.1041**
	(0.0503)	(0.0476)	(0.0460)
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	187,806	187,806	187,806
R^2	0.7616	0.8779	0.9359

The table reports the results of regressing the natural logarithm of district level median 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:

$$Ln(1 + Y_{i,j,t}) = \beta \cdot \text{Direct Rule}_i + \theta_{j,p(i \in p),t} + Latitude_j + Longitude_j \varepsilon_{i,j,t}$$

where i denotes a district with a median 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. Standard errors reported in parentheses are two-way clustered by native states and district-pair. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Investment Share and Settler Mortality

Dep Var: Investment Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Mortality)	-0.0097***	-0.0098***	-0.0096***	-0.0069***	-0.0100***	-0.0043*	-0.0036**
	(0.0017)	(0.0017)	(0.0017)	(0.0019)	(0.0017)	(0.0022)	(0.0016)
Latitude		-0.0319				-0.0171	0.0035
		(0.0192)				(0.0162)	(0.0037)
Forced Labor			-0.0032			0.0014	-0.0027
			(0.0026)			(0.0038)	(0.0029)
ln(Land Area)				-0.0015		-0.0046***	-0.0034***
				(0.0010)		(0.0017)	(0.0009)
ln(Population)					0.0006	0.0055**	0.0042***
					(0.0023)	(0.0026)	(0.0013)
Constant	0.0519***	0.0573***	0.0518***	0.0549***	0.0438	-0.0132	
	(0.0073)	(0.0056)	(0.0073)	(0.0078)	(0.0356)	(0.0356)	
Region FE							Yes
# Obs	63	63	63	63	63	63	63
R^2	0.5418	0.6229	0.5421	0.5761	0.5449	0.7214	0.9572

The table reports the results of regressing the share of direct investment received by a country from G20 countries between 2009 and 2019, on settler mortality, à la Acemoglu, Johnson and Robinson (2001). The specification includes controls for latitude, forced labor (slavery), natural logarithm of geographic area, and natural log of total population. We also control for world-bank region fixed effects in column (7). Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Investment Share and Governance Indicators

Dep Var: Investment Share	(1)	(2)	(3)	(4)
Government Effectiveness	0.0112***			
	(0.0034)			
Regulatory Quality		0.0110***		
		(0.0031)		
Rule of Law			0.0118***	
			(0.0034)	
Control of Corruption				0.0114***
				(0.0034)
Origin Country × Year FE	Yes	Yes	Yes	Yes
Destination Region × Year	Yes	Yes	Yes	Yes
# Obs	23,784	23,784	23,837	23,784
R^2	0.2025	0.2008	0.1985	0.2041

The table reports the results of regressing the share of direct investment received by a country from G20 countries on different measures of institutional quality between 2009 and 2019. Each observation in data represents a country pair – destination and origin country. The sample of destination countries include all countries. The sample of origin (investing) countries include G20 countries. The four measures of institutional quality come from the World Bank Governance Indicators and includes government effectiveness, regulatory quality, rule of law and control for corruption. All specifications include origin country × year fixed effects and destination region × year fixed effects. Regions are defined using the World Bank classification of regions. Standard errors reported in parentheses are two-way clustered by origin and destination countries. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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.[13] 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

[13]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.

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 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 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 male 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.[14] 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,

[14]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.

over thousands of square miles. Some princely states had maximum administrative powers while other princely states had minimal administrative powers.[15] 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 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.

[15] 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.

Appendix B Data

B.1 Project Announcement - CapEx Database

Table B.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 B.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 B.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 C 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:

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

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.

C.1 Construction of Weight Matrix

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

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

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

C.1.1 Inverse Weights

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

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

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.

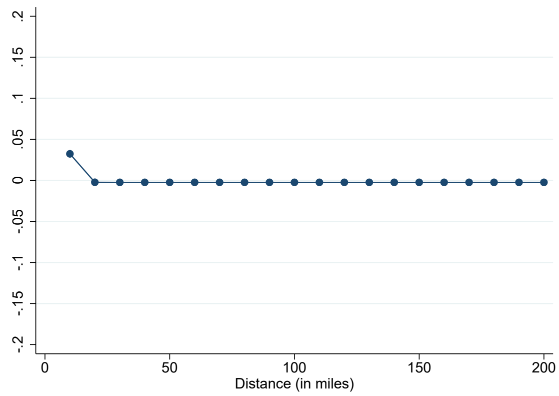
C.1.2 Uniform Weights

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

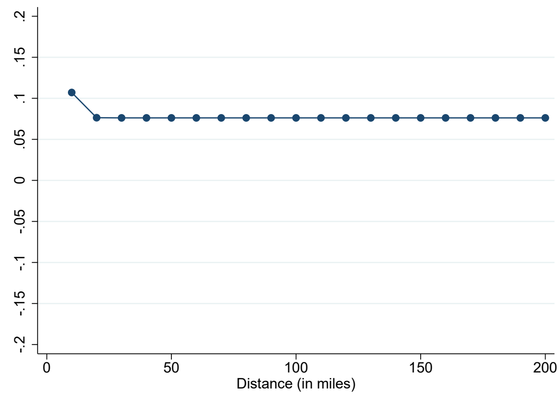
$$\begin{cases} \frac{\mathbb{1}_{d_{ij}<d}}{\sum_{j=1}^{n} \mathbb{1}_{d_{ij}<d}} & d_{ij} < d, i \neq j \\ 0 & otherwise \end{cases}$$

where $\mathbb{1}_{d_{ij}<d}$ 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 C.1: Spatial Autocorrelation - Moran I statistic



(a) Uniformly Weighted

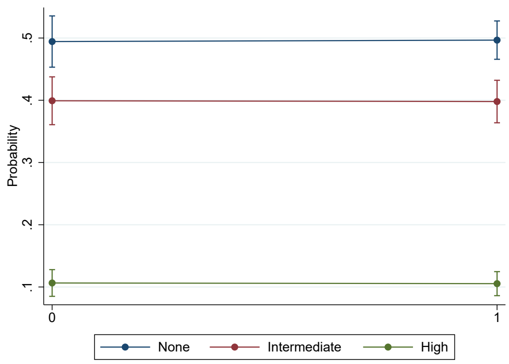


(b) Inverse Distance Weights

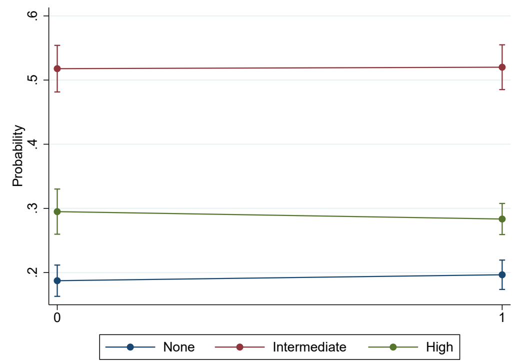
The figure plots the Moran I statistic for the Direct British Rule variable. The Moran I is computed for different distance thresholds from 10 miles to 200 miles. Figure C.1a uses the uniform weight between two points if they are within the distance limit and 0 otherwise. The weight takes a value of 1 if the two points are within the threshold and 0 otherwise. Figure C.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 D   Alternative Explanations:

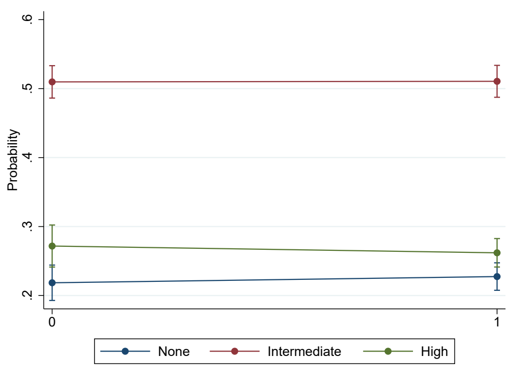
Figure D.1: Trust in State Apparatus



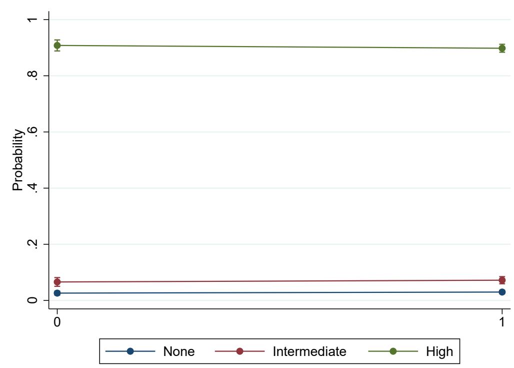
(a) Trust in Politicians



(b) Trust in Panchayats



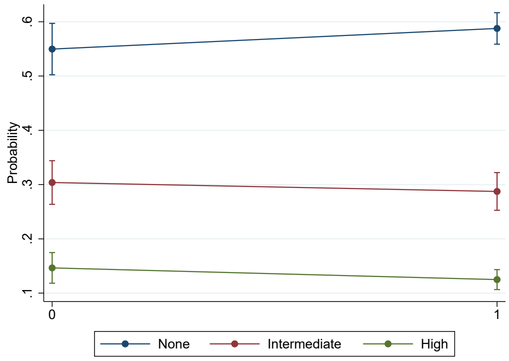
(c) Trust in Police



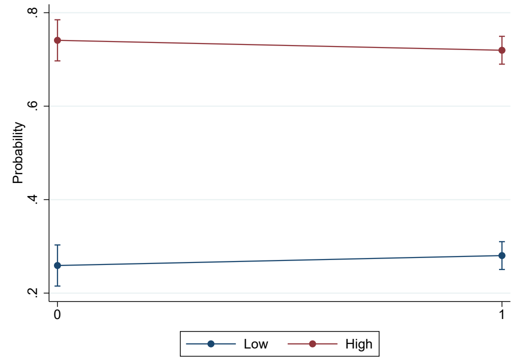
(d) Trust in Banks

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 D.2: 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 D.2a. 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 D.2b. 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.