CESifo WORKING PAPERS

9031
2021

April 2021

Impact of Colonial Institutions on Economic Growth and Development in India: Evidence from Night Lights Data

Priyaranjan Jha, Karan Talathi

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9031*

Impact of Colonial Institutions on Economic Growth and Development in India: Evidence from Night Lights Data

Abstract

We study the implications of two historical institutions, direct British rule, and the heterogeneous land tenure institutions implemented by the British, on disparity in present day development using district level data from India. Using nightlights per capita as a proxy for district level per capita income, we find that modern districts that were historically under direct British rule had 39.47% less nightlights per capita in 1993 relative to modern districts that were historically under indirect British rule. The large gap persists even after including other controls such as educational attainment, health, and physical infrastructure. Looking at the growth pattern during 1993 to 2013, directly ruled districts had a 1.84% lower annual growth rate compared to indirectly ruled districts. As well, directly ruled districts were converging at a rate of 2% per year while indirectly ruled districts were converging at a rate of 5.7% per year. Much of the development gap between areas under indirect rule and direct rule can be accounted for by the adverse effect of landlord-based revenue collection system in the directly ruled areas.

JEL-Codes: O110, O430, P160, P510.

Keywords: institutions, direct British rule, economic growth, nightlights per capita, land tenure system, economic development, human capital.

Priyaranjan Jha
University of California, Irvine / USA
pranjan@uci.edu

Karan Talathi
University of California, Irvine / USA
ktalathi@uci.edu

April 2021
We thank Dan Bogart for his insightful comments throughout the development of this paper. We also thank Areendam Chanda for clarifying important aspects of GIS data and comments on the paper. We have also benefited from comments from Tejaswi Velayudhan, Kamal Sadiq, Antonio Rodriguez-Lopez, and members of the THPE workshop at University of California, Irvine. All errors are our own.

I. Introduction

Many cross-country studies find that quality of historical institutions is a major cause of disparity in present-day economic development as measured by income per capita.[1] Due to the unavailability of data on a comprehensive measure of development such as per capita income, studies examining the role of historical institutions on development at the sub-national levels use alternate proxies of economic well-being in their analysis.[2] In this paper, we examine the long-term effects of British colonial institutions on overall economic development within India using satellite nightlights data.

During the period of British rule India was divided into two types of territories: British India and Native or Princely States. British India comprised of areas where the British administration had full autonomy in the internal and external affairs and hence were under the 'direct rule' of the British. Princely states, on the other hand, were areas that were ruled by the local kings (or hereditary rulers) and therefore, were under the 'indirect rule' of the British. While the external affairs of the princely states were under the British control, the local kings (or hereditary rulers) had full autonomy in the internal affairs of these areas. This interesting characteristic of the native states makes these regions a good counterfactual to the areas that were directly governed by the British colonial rule. After the Independence of India in 1947, all these regions collectively came under a uniform governance. We exploit this division of India into areas that came directly under British rule and areas that were indirectly governed by the British rule during the colonial period and investigate the effect of being directly ruled by the British on present-day economic development.

The effect of this colonial institution was first studied by Iyer (2010). Her key finding is that after controlling for the selection effect, the areas under direct rule by the British had lower levels of investment in public goods such as health, education, canals, and roads. She also finds directly ruled areas to have higher levels of poverty, inequality, infant mortality but similar levels of literacy. Iyer (2010) notes that "One major drawback of district-level data in India is the absence of data on per capita income, consumption, or net domestic product (these are available only at the state level)", as a result her study is unable to study the impact of this historical institution on

[1] See Acemoglu et al. (2001), Rodrik et al. (2004), Nunn (2008), Feyrer and Sacerdote (2009).

[2] See Dell (2010), Dell and Olken (2019), Banerjee and Iyer (2005), Banerjee et al. (2005), Iyer (2010).

overall development. We extend Iyer's analysis by constructing a proxy for overall economic development at the district-level in India using satellite nightlights data. The satellite nightlights data has many advantages including its availability at higher levels of spatial disaggregation and has been widely accepted as a proxy for overall economic development at the national and sub-national levels in the growth and development literature.[3]

From our benchmark OLS specifications, we find that modern districts that were historically under direct British rule had 39.47% less nightlights per capita in 1993 relative to modern districts that were historically under indirect British rule. Looking at the growth pattern during 1993 to 2013, a period of rapid growth following the liberalization that began in 1991, we find that areas that were under direct British rule had a 1.84% lower annual growth rate compared to indirectly ruled areas. The negative coefficient of the initial level of nightlights per capita provides evidence of convergence, that is, areas that were initially less developed were growing faster. On analyzing the rate of convergence across areas under direct and indirect rule, we find that areas under direct British rule were converging at a rate of 2% per year while areas that were under indirect British rule were converging at a rate of 5.7% per year.

Our OLS results could be subject to the selection bias if the British selectively annexed areas that came under their direct rule. For example, if they annexed areas with low development potential (they were weaker to put up too much fight), this could drive a negative relationship between direct British rule and nightlights per capita without a causal relationship. As a first check on the selection issue, we restrict our sample to only those directly ruled districts which have an adjacent indirectly ruled district. Adjacent districts are not likely to differ much in their development potential. Even in this restricted sample of neighboring districts, directly ruled districts have a significantly lower nightlights per capita.

Iyer (2010) provides convincing evidence that the British selectively annexed areas that had higher agricultural productivity. Using an instrument for direct British rule, based on the policy of *Doctrine of Lapse* adopted by the British from 1848-1856 whereby the death of a ruler of a princely state without an heir would be automatically annexed, Iyer (2010) shows that direct British rule was much more damaging in terms of public investment in physical and human infrastructure than what is captured by the OLS estimates. Using the same instrument as Iyer

[3] Henderson et al. (2012), Chen and Nordhaus (2011), Donaldson and Storeygard (2016), Pinkovskiy and Sala-i-Martin (2016), Chanda and Kabiraj (2020), Prakash, Rockmore, and Uppal (2019)

(2010), we find that directly ruled modern districts had 46.69 – 48.88% (depending on the specification) less nightlights per capita in 1993 relative to indirectly ruled modern districts. Our finding that directly ruled districts are doing worse when correcting for the selection effect is consistent with the story of the British annexing more productive areas which would make the OLS results underestimate the adverse effect of direct British rule.

Next, we turn to the possible channels through which direct British rule may account for the relative backwardness of these districts compared to the indirectly ruled districts. Given the importance of human capital in the development process, and the debate on the primacy of institutions vs human capital[4], we include the level of human capital measured by literacy rate as an additional regressor. While the literacy rate is strongly positively related with nightlights per capita, the coefficient of direct British rule increases upon the inclusion of literacy rate. This is explained by the slight positive correlation between direct British rule and literacy rate. Therefore, inadequate human capital due to direct British rule cannot explain the relative backwardness of these areas.[5]

Iyer (2010) also found evidence of underinvestment in health and physical infrastructure in directly ruled areas. When we include infant mortality rate as a measure of health, and roads per capita and railroads per capita to capture physical infrastructure, we still find a substantial negative effect of direct British rule on nightlights per capita. That is, some of the adverse effects of direct British rule may be occurring through lower health and worse roads (British districts had more railroads per capita), but a substantial part remains unexplained.

In trying to understand the mechanisms through which direct British rule may have adversely affected development, Iyer (2010) finds a role for the land revenue collection systems (land tenure systems) developed by the British. She finds that within the areas under direct British rule, districts under non-landlord-based revenue collection system had more investment in physical and human infrastructure than those under landlord-based revenue collection system. No such difference existed in areas under indirect rule.

Studying the implications of different land tenure systems in the colonial period on nightlights per capita, we find that a large part of the difference in overall development between

[4] See Glaeser et al. (2004) and Acemoglu et al. (2014).

[5] In our robustness exercise we also use some alternative measures of human capital such as the proportion of population above 25 with completed secondary schooling and obtain similar results.

directly and indirectly ruled districts is driven by directly ruled landlord districts doing much worse. The difference between directly and indirectly ruled districts with a non-landlord-based revenue collection system is much smaller. Also, the differential revenue collection institutions do not affect overall development across indirectly ruled districts. More importantly, these results persist even after controlling for literacy, infant mortality, and measures of physical infrastructure. Restricting the sample to only those landlord districts which have a neighboring non-landlord district yields similar results.

In sum, we have documented a significant difference in the levels of development between districts under direct British rule and indirect rule. As well, direct British rule in combination with the landlord-based revenue system has a persistent effect on development through channels other than or in addition to health, human capital, and infrastructure.

What explains the pernicious effects of landlord-based system in directly ruled areas? Banerjee and Iyer (2005) was the first study to document the adverse effects of landlord-based system on agricultural investments and public good provision, though only within the directly ruled districts. They conjectured that the landlord-based system created an entrenched elite giving rise to class-based divide and conflict within the society making collective action in the provision of public goods more difficult. Pandey (2010) provides corroborating microlevel evidence by comparing the governance and educational outcomes in landlord and non-landlord districts in Uttar Pradesh (the most populous state in India). She finds that villages in non-landlord districts did much better in terms of both governance and educational outcomes. Since the literacy rate and school enrollment rates were similar across non-landlord and landlord districts, her work suggests that quality of human capital may be different. For a subset of districts in our sample we were able to obtain test scores and we used it as a proxy for the quality of human capital. Test scores are slightly higher in the directly ruled districts compared to indirectly ruled districts. They are also slightly higher in non-landlord districts, however, the coefficient of non-landlord dummy fails to be statistically significant. Therefore, the results based on the limited data that we have do not provide a clear-cut evidence that the quality of education is worse in landlord districts. When we include test scores as a measure of human capital instead of literacy rate in our regressions, the results are qualitatively similar to those obtained using the literacy rate. That is, nightlights per capita are significantly lower in directly ruled landlord districts compared to the other 3 types of districts.

Next, we perform a series of checks to ensure the robustness of our results. We first account for issues prevalent in the measurement of nightlights data such as top-coding and bottom censoring as well as overglow and blooming effect of nightlights that could bias our results.[6] For this, we use an alternative nightlights dataset by Bluhm and Krause (2018) who correct the 'stable' nightlights images for top-coding and bottom censoring. We also control for the distance to the closest major city from the centroid of each district to account for the overglow and blooming effect of nightlights. Our results are robust to these measurement issues in nightlights data. Secondly, since our analysis is limited to a cross-section and focuses on outcomes in 1993, it is possible that our results may vary with outcomes from different years. We check for this by using nightlights per capita in 2013, 2016, and 2019 (instead of 1993) as the outcome variable. We find that the adverse effect of direct British rule on present-day overall development has persisted till 2019. Interestingly, while directly ruled districts are still doing worse, the gap has narrowed down in 2019 compared to 1993. An additional concern of our analysis is that the exclusion of districts from smaller states, a common practice in the literature, may bias our main results. When we extend the analysis to include districts from smaller states the results remain robust. Fourthly, our use of a dummy variable to represent if a modern district (1991-level) in India was historically under the direct rule of the British could potentially be a cause of worry as it omits the differences in effect of direct rule stemming from differences in time spent under the direct British rule.[7] We use the duration spent under British rule as our main explanatory variable instead of the direct British rule dummy and find that our results are unchanged. As a fifth robustness check, we account for differences in the measurement of nightlights, agricultural suitability, and initial differences across modern districts. For this, we include additional controls specifically terrain ruggedness, agricultural suitability index, and historical controls and find that our results are robust to the inclusion of these additional controls. Lastly, we use an alternative proxy of development, specifically consumption per capita data for rural districts only, instead of nightlights per capita and find that our results remain highly robust.[8]

Among other related papers, Castello-Climent et al. (2018) study the implications of human capital for development using district level nightlights data from India. To tackle the endogeneity

---

[6] Pinkovskiy and Sala-i-Martin (2016), Henderson et al. (2012), Chanda and Kabiraj (2020), Prakash, Rockmore, and Uppal (2019)

[7] Feyrer and Sacerdote (2009)

[8] The per capita consumption data at the district level are available only for rural districts.

of human capital, they use the historical location of Catholic missionaries in India in 1911 as an instrument for present-day human capital accumulation.[9] They find a large and significant positive effect of human capital on the density of lights across districts in India. They also include a dummy for whether a district was part of a princely state (indirect British rule) during the colonial period and find this dummy to be insignificant. Since we use nightlights per capita as our dependent variable rather than nightlights density[10], the results are not directly comparable. However, since human capital is a possible mechanism through which institutions can affect development, we not only control for human capital, but also use their instrument, the location of Catholic missionaries. In fact, along the lines of Acemoglu et al. (2014) where they instrument for both institutions and human capital, we also provide estimates where we instrument for both direct British rule and literacy rate (our preferred measure of human capital) and find that direct British rule has a large negative effect and literacy has a large positive effect on development measured by nightlights per capita.

Compared to other studies looking at the impact of the direct British rule and/or colonial land tenure institutions in India, we make 4 contributions: 1) Using a comprehensive measure of development, nightlights per capita, we show that the areas under direct British rule had a much lower level of development compared to areas under indirect rule; 2) Conditional on initial level of development, areas under direct British rule grew at a much slower pace during the period 1993-2013; 3) Much of the development gap between areas under indirect rule and direct rule can be accounted for by the adverse effect of landlord-based revenue collection system in the directly ruled areas; 4) The development gap remains even after controlling for education, health, and physical infrastructure.

Our study contributes to the literature examining the persistent effects of historical institutions on present-day economic development across smaller geographical units within nations. Among studies not related to India, Dell (2010) studies the long run effects of *mita*, an extensive forced mining labor system operational in Peru and Bolivia between 1573 and 1812. Using data from Peru, she finds that in *mita* districts household consumption was lower and the

---

[9] Woodberry (2004), Gallego and Woodberry (2009, 2010), Becker and Woessmann (2009), Woodberry (2011), Acemoglu et al. (2014) are other studies using the location of Catholic/Protestant missionaries to instrument human capital.

[10] In the robustness exercise in Table 11, they do report results with lights per capita, but the coefficient of Princely state is not reported. Also, they use 2001 district boundaries and have 500 districts in their analysis which makes the comparison with our study or Iyer (2010) that use 1991 district boundaries difficult. In any case, we confirm the robustness of our results using nightlights density.

incidence of stunting higher compared to adjacent districts that were exempt from *mita*. Dell and Olken (2019) study the long-run impact of the Dutch Cultivation System in operation in the 19th century Java for the production of sugar on present-day outcomes. They find that areas close to the location of sugar factories established by the Dutch in the mid-19th century are doing better in terms of infrastructure, industrialization, education, public goods provision, and household consumption compared to similar areas which did not get a factory.

The remainder of our paper is structured as follows. Section II provides details of the data we use in our analysis. In Section III, we describe our empirical specifications and discuss our main results. We perform a series of checks to ensure the robustness of our main results in Section IV. Section V provides some concluding remarks.

II. Data

In this section, we describe the main variables we use in our empirical exercise. The details on each variable and their source are provided in the Data appendix. Table 1 provides the summary statistics for the variables we use in our analysis.

II.1. Nightlights Per Capita

The nightlights data are obtained from the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC). The nightlights data is in the form of images of lights generated from the earth's surface. These images are available for years spanning 1992-2013.[11] We use the *sum of lights* statistic that represents the total luminosity of nightlights emitted in 1993 from each district and divide it by the district population to obtain *Nightlights Per Capita in 1993* which is our proxy for per capita income at the district level.[12] Figure 1 showcases the raw nightlights image for the year 1993 for India.

[11] The NOAA makes multiple nightlights data products available publicly. In our main specification, we use the 'stable' nightlights product of the DMP-OLS nightlights data series. This data is top-coded at 63. We deal with this issue as well as other measurement issues of the nightlights data in our robustness analysis in section IV.

[12] Since district-level population is not available for 1993, we use the 1991 district-level population to compute *Nightlights Per Capita in 1993.*

II.2. Institutional Data

The data on districts that were historically under direct British rule and districts that were historically a part of a Princely state are obtained from Iyer (2010).[13] Figure 2 depicts the distribution of the natural log of nightlights per capita in 1993 across the 466 districts in India. The directly ruled districts are in blue and the indirectly ruled districts are in red. Darker shades capture more lights.

Though India was divided into 466 districts as per the 1991 district boundaries, we focus on 412 districts spread across 18 major states of India (1991 Census). Focusing on major states within India is a common practice and involves dropping observations from small States, Union Territories (UTs), and the North-Eastern States for several reasons including the quality of current data, problems in matching current and historic district boundaries, and availability of data.[14] Out of these 412 modern (1991-level) districts, 265 districts were historically under the direct British rule while 147 districts were historically a part of princely or native states.

III. Empirical Specification and Main Results

III.1. Empirical Specification: Baseline OLS level regressions

We first identify the long-term effects of the British colonial legacy using a simple cross-sectional OLS model described in equation (1).

$$y_{i,j} = \alpha_0 + \beta Brit_{i,j} + \gamma X_{i,j} + \delta_j + \varepsilon_{i,j} \quad \ldots\ldots\ldots\ldots\ldots.. (1)$$

[13] The historical background for why a district in India was under direct British rule or a Princely state (under indirect rule) is discussed in detail in Iyer (2010). Also, as described in detail in the data appendix and Table A10 in the online appendix, compared to Iyer (2010) we switched classifications for 5 districts based on the evidence provided in two recent papers, Verghese (2019) and Castello-Climent et al. (2018), and our own reading of the evidence on this issue. The slight difference in the classification of modern districts into directly ruled or indirectly ruled in various studies arises because some districts contain areas from both British India and Princely states.

[14] The 18 major states we focus on are: Andhra Pradesh, Assam, Bihar, Delhi, Gujarat, Haryana, Himachal Pradesh, Jammu and Kashmir, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal. Iyer (2010) conducts her analysis for 17 major states of India which excludes Delhi from our 18 states. Banerjee and Iyer (2005) conduct their analysis for 13 major states, Besley and Burgess (2000) for 16 major states, and Castello-Climent et al. (2018) for 20 major states of India.

We use the natural log of nightlights per capita in 1993 for district i in state j as our main outcome variable which is represented by $y_{i,j}$ in equation (1).[15] Our main independent variable, $Brit_{i,j}$ takes the value of 1 if the district was under direct British rule and 0 otherwise. $X_{i,j}$ represents the set of geographical controls used to capture some of the heterogeneity across the districts as well as to account for the direct impact of geography on development. Following Pinkovskiy and Sala-i-martin (2016) who also use nightlights per capita as the dependent variable, we control for the area of each district in 1991. This is mainly to account for potentially higher luminosity emission from smaller areas due to the higher population densities. Our other geographical controls are average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast for each district to account for heterogeneity in geographical characteristics that leads to differences in agricultural productivity and eventually differences in development across the districts. These geographical characteristics also affect climatic conditions which further affect measurement errors in nightlights data. In addition to its effect on agricultural productivity and climate, distance to the nearest coast from the centroid of each district can also affect development through the trade channel as areas closer to the coast are more likely to be involved in trade and hence may have higher levels of economic development. In our results tables, we refer to the set of these controls as 'Geographical Controls'.

δ_j in equation (1) represents an indicator for state j of district i. The state here refers to the current Indian state that the district belongs to. We include these state-fixed effects in some of our regressions to account for the heterogeneity in state-level policies and administrative decisions that could impact development outcomes of districts spread across heterogenous modern states. In all our regressions, we cluster standard errors at the native state level to account for potential correlation in outcomes across modern districts that historically were a part of the same native state.[16] The native state here refers to the Princely state that the district belonged to during the British rule.

[15] Though the earliest available nightlights data is for the year 1992, we use the nightlights data from 1993 in our analysis. The main reason for this is the quality of the nightlights data. The 1992 district-level nightlights data for India has more observations with 0 *sum of lights* relative to the 1993 data which results in larger spikes in the data and increased noise in our regressions. About 3% of the entire 1992 nightlights data for India has *sum of lights* equal to 0 relative to the 1% in the 1993 nightlights data for India.

[16] And, following Iyer (2010), for districts under direct British rule, the clustering is done according to region and date of annexation.

The role of institutions as the key determinant of economic growth and development is widely accepted. An alternative view, originally ascribed to Lipset (1959), posits that improvement in institutional quality is a result of growth in income and human capital accumulation. Supporting this view, Glaeser et al. (2004) criticize the seminal study by Acemoglu et al. (2001) for not accounting for the role of human capital in their analysis. Glaeser et al. (2004) show that, after controlling for human capital accumulation, institutions no longer have a significant impact on long-run development. Responding to this criticism, Acemoglu et al. (2014) show that the impact of institutions on long-run development is robust and the significance of human capital in explaining the differences in per capita income across countries goes down when both institutions and human capital are instrumented. Chanda et al. (2014), in their persistence of fortune study, find results consistent with Glaeser et al. (2014). Since we are studying the implications of a historical institution on economic development, following this debate, we control for human capital in equation (1) above. Our measure of human capital is the literacy rate in 1991 but we also perform robustness checks with other measures such as the share of population with completed secondary schooling.

III.2. Baseline OLS results on the level of development

Table 2 reports the baseline OLS results. In column 1, we report the OLS result obtained from the specification described in equation (1). The coefficient on $Brit_{i,j}$ in column 1 is negative and statistically significant and implies that modern districts that were under direct rule of the British have 39.47% less lights per capita in 1993 compared to the districts that were under Princely states.[17] In column 2, we include $Lit.Rate_{i,j}$ as an additional regressor. The positive and significant coefficient on $Lit.Rate_{i,j}$ suggests that an increase in literacy rate by 1 percentage point (pp hereon) leads to a 4.41% increase in nightlights per capita. The coefficient on $Brit_{i,j}$ remains negative and statistically significant. In fact, the negative coefficient on $Brit_{i,j}$ increases in magnitude in column 2 indicating a positive correlation between $Brit_{i,j}$ and $Lit.Rate_{i,j}$, which we confirm in the data as well. This result is somewhat at odds with the findings in Iyer (2010) that indirectly ruled districts have greater investment in human capital in the form of number of villages

[17] Alternatively, districts under indirect rule had 65% more lights per capita.

having primary, middle, and high schools. Higher investments in human capital do not seem to translate into higher literacy rates in indirectly ruled districts, something that was found by Iyer (2010) as well. More generally, the findings suggest that the direct British rule's adverse effect on long-run development occurs through channels other than human capital.

Having looked at the effect of direct British rule on the level of development in 1993, and before turning to the issues of selection bias, omitted variable bias etc. in our OLS estimates, we next analyze if the direct British rule has any growth effects during the period for which we have nightlights data. This is an important question because India embarked on a policy of massive liberalization, both internal and external, starting in 1991 which resulted in a rapid growth in per capita income over the next couple of decades.

III.3. Impact on Growth and Convergence

To examine if the British colonial legacy affected differentially the growth rates of districts that were under direct and indirect rules, we estimate the specification described in the equation below.

$$growth_{i,j,t,t-k} = \alpha_4 + \rho y_{i,j,t-k} + \beta Brit_{i,j} + \phi Lit.Rate_{i,j,t-k} + \varepsilon_{i,j} \quad \ldots\ldots\ldots\ldots\ldots.. (2)$$

In equation (2), $growth_{i,j,t,t-k}$ represents the growth in nightlights per capita of district *i* in state *j* between years *t* (2013) and *t-k* (1993). $y_{i,j,t-k}$ denotes nightlights per capita in 1993 which is included as a regressor to test for convergence across districts following the prediction of the neoclassical growth model that countries/regions that are poorer tend to grow faster. The inclusion of $Brit_{i,j}$ and $Lit.Rate_{i,j}$ allow for conditional convergence as we describe below.

Table 3 reports the regression results of the specification described in equation (2). In column 1, we only include $y_{i,j,t-k}$. The coefficient on $y_{i,j,t-k}$ is -0.0231 which is very close to the convergence coefficient of -0.02 found in Barro and Sala-i-Martin (1992) for US states from 1880 to 1988. This result is also in line with Chanda and Kabiraj (2020) who also use nightlights data and provide evidence across districts within India from 1996-2010. The convergence coefficient of -0.0231 implies that districts across India are converging at a rate of 3.1% per year and it will take 30 years to close half the gap.[18] In column 2, we add $Brit_{i,j}$ as an explanatory variable. The coefficient on $Brit_{i,j}$ is negative and significant suggesting that after accounting for heterogeneity in their initial levels of nightlights per capita in 1993, modern districts that were under direct British rule experienced a 1.84% lower growth rate annually on average as compared to districts that were under indirect British rule. Adding $Lit.Rate_{i,j}$ as an explanatory variable in column 3 we find that the districts with higher literacy in 1993 grew faster. The coefficient of $Brit_{i,j}$ remains negative and significant. Columns 2 and 3 provide evidence for conditional convergence, that is, districts in both groups, under direct and indirect rule, converge to their respective steady states. In columns 4 and 6, we estimate convergence regressions separately for directly and indirectly ruled districts. Both types of districts converge to their respective steady states as evidenced by the negative and significant coefficients of $y_{i,j,t-k}$ in columns 4 and 6. We also note that the districts under direct British rule (column 4) converge at a slower rate of 2% per year than the districts under indirect rule (column 6) which converge at a rate of 5.7% per year. When we add the initial literacy rate in columns 5 and 7, the results remain qualitatively similar. While conditioning on literacy increases the pace of convergence among directly ruled British districts (the coefficient increases in magnitude from .0165 to .206), the convergence coefficient is unchanged for indirectly ruled districts.

Overall, we learn that in addition to negatively affecting present-day overall development, direct British rule also adversely affected present-day growth from 1993 to 2013.

Therefore, both our level and growth regressions suggest the adverse impact of direct British rule on overall development as measured by nightlights per capita.

In the remainder of the paper, we focus on level regressions to establish the robustness of OLS results and to explore the mechanisms through which direct British rule may have a persistent effect on development.

III.4. Neighborhood district analysis

It is conceivable that some omitted factors may explain the negative association between development and directly ruled districts. One way to reduce the influence of omitted factors is to

[18] Convergence rate is calculated using $\rho=-(1-e^{-\pi T})/T$, where π is the convergence rate and ρ is the convergence coefficient. Half-life is calculated using $h=\ln(0.5)/\rho$, where h is number of years taken to close half the gap.

restrict the sample to only neighboring districts. That is, we restrict the sample to only those directly ruled districts which have a bordering indirectly ruled district. This reduces the number of districts to 221 but other omitted factors are likely to be similar in these neighboring districts. Figure 3 shows lights per capita in the neighboring districts sample. The regression results are reported in columns 3 and 4 of Table 2. The coefficient of $Brit_{i,j}$ is negative but fails to be statistically significant in column 2. However, quantitatively it is still significant implying that indirectly ruled districts have 17% more lights per capita than directly ruled districts. When we include $Lit.Rate_{i,j}$ as an additional control in column 4, the coefficient of $Lit.Rate_{i,j}$ is positive and significant as was the case in column 2 for all districts. Also, the inclusion of $Lit.Rate_{i,j}$ increases the magnitude of the coefficient of $Brit_{i,j}$ in addition to making it statistically significant. Therefore, our neighboring district analysis confirms the validity of our baseline OLS results.

III.5. IV Regressions

III.5.1. Endogeneity of $Brit_{i,j}$ stemming from selective annexation:

In the OLS specification described in equation (1), the assumption is that annexation of areas by the British was random. However, Iyer (2010) finds evidence that the British annexation policy was selective towards areas more favorable for agriculture. Our geographical controls already alleviate this concern, but we use an instrumental variable approach to account for the underlying endogeneity of $Brit_{i,j}$ stemming from selective annexation. We use the instrument, *Lapse*, first constructed and used by Iyer (2010) who also provides a detailed description of the rationale behind using this instrument. Here is a brief description of it. The British used a specific annexation policy, called the *Doctrine of Lapse*, from 1848-1856 under which if the hereditary ruler of a princely state died without a natural heir, the state would automatically be annexed by the British i.e., it would lose its princely status and come under the direct rule of the British. The death of a ruler without a natural heir is a random event and is unlikely to affect development outcomes in the post-colonial period. According to this annexation policy, a princely state whose hereditary ruler died without a natural heir was relatively more likely to be annexed by the British between 1848-1856. Therefore, our instrument for direct British rule, $Brit_{i,j}$, is a dummy variable $Lapse_{i,j}$ for each modern 1991 district i in state j which takes the value 1 if the district was part of a princely state that was not annexed before 1848 and the hereditary ruler of that princely state died without a natural heir between 1848-1856 and a value of 0 if the princely state was not annexed before 1848 and the hereditary ruler of that state did not die without a natural heir between 1848-1856.

Similar to Iyer (2010), in our IV analysis we only consider areas that were either never annexed throughout the colonial rule in India or were annexed by the British between 1848-1856. The total sample size is reduced to 181 districts out of which 143 were historically part of princely states that were never annexed and the hereditary rulers of these princely states did not die without a natural heir between 1848-1856. 4 districts were historically part of princely states whose hereditary rulers died without a natural heir between 1848-1856 but were never annexed by the British due to extraneous factors while 19 districts were historically part of princely states whose hereditary rulers did not die without a natural heir between 1848-1856 but were annexed by the British through other means. The remaining 15 districts were historically part of native states whose rulers died without natural heirs between 1848-1856 and consequentially were brought under the direct rule of the British through the *Doctrine of Lapse*.

III.5.2. Endogeneity of Literacy rate:

Since our main interest lies in ascertaining the importance of direct British rule on development, the endogeneity of literacy rate (reverse causality from nightlights to literacy) which is included as a key regressor in our baseline regression is not a prime concern. However, it also gives us an opportunity to run a horse-race regression between institutions and human capital as the alternative determinants of development where both these endogenous variables are instrumented as in Acemoglu et al. (2014). We follow the method of Woodberry (2004), Gallego and Woodberry (2009, 2010), Becker and Woessman (2009), Woodberry (2011), Acemoglu et al. (2014), and Castello-Climent et al. (2018) and use the historical location of Catholic missionaries in the early 20th century as an instrument for present-day literacy rate. The precise instrument that we use in the Indian context is the one used by Castello-Climent et al. (2018) who find that historical location of Catholic missionaries is highly correlated with present-day human capital accumulation across districts in India.[19]

Having described our instruments, we now discuss our main results in the next section.

[19] For more information on data for the historical location of Catholic missionaries in 1911, refer to the data appendix.

III.6. IV Regression Results

IV regression results for our baseline estimating equation are presented in Table 2. As mentioned earlier, our IV sample has only 181 districts while the OLS sample has 412 districts. To keep the IV results comparable to the OLS results, in column 5 we provide the OLS results for the IV sub-sample of 181 districts and they are similar to the OLS results for the full sample presented in column 2. The IV results when $Brit_{i,j}$, is instrumented by $Lapse_{i,j}$ is shown in column 6. The corresponding first stage results are reported in the same column in panel B of Table 2. The coefficient on $Lapse_{i,j}$ in the first stage is positive and significant implying that if the hereditary ruler of a princely state died without a natural heir between 1848-1856, the probability that it was automatically annexed to the British empire through the *Doctrine of Lapse* was 59.5%.[20] The coefficient on $Brit_{i,j}$ in column 6, is negative and significant and close to the OLS coefficient in column 5. More interestingly, the coefficient of $Brit_{i,j}$ in column 6 is larger than the coefficient for the full sample in column 2 suggesting that the selective annexation in the pre-1848 period is causing a downward bias in our OLS estimates. This is consistent with the finding in Iyer (2010) that in the pre-1848 period the British annexation policy was selective towards areas more favorable for agriculture. Since agriculture accounted for a significant proportion of GDP in 1993, we would expect the agricultural GDP per capita to be higher in British districts annexed before 1848, and this factor by itself would create a positive relationship between direct British rule and the level of development. Therefore, our OLS finding of a negative relationship between our measure of development and the dummy for direct British rule is likely to be biased downwards due to the selection effect. However, we should also mention that the IV estimate could be biased upwards due to measurement error issues.

In columns 7 and 8 of Table 2, we account for the endogeneity of the literacy rate by instrumenting it with the historical location of Catholic missionaries. Not instrumenting $Brit_{i,j}$ in column 7 allows us to see the impact of instrumenting literacy for the whole sample of 412 districts. This makes our estimates in column 7 comparable to the estimates in Castello-Climent et al. (2018) whose main interest is in studying the impact of human capital on development where they also use nightlights as a proxy for district level development. The first stage results reported in panel C

[20] The Kleibergen and Paap F statistics for the first stage is 10.399 which lies above the Stock and Yogo (2005) critical value for 15% maximal IV size of 8.96. Hence weak identification is not an issue.

show that the literacy rate in 1991 is 5.9 pp greater on average in districts that had a Catholic missionary in 1911 relative to districts that did not have one.[21] The coefficient of literacy in the second stage is positive and significant which is in line with the results in Castello-Climent et al. (2018). They also include a dummy for Princely states (indirect rule) and find it to be insignificantly related with nightlights. In contrast, the coefficient of $Brit_{i,j}$ in column 7 is large and significant suggesting that directly ruled British districts have 47.59% less nightlights per capita relative to indirectly ruled districts. Castello-Climent et al. (2018) use nightlights density per unit of area instead of nightlights per capita as their dependent variable in their main specifications. Their measure of human capital is also different. They use the share of population above 25 with higher education and separately the share of population above 25 with primary and middle education. In our online appendix tables A1-A4 we use nighlights density as the dependent variable and their measure of human capital and obtain results similar to those reported in column 7 in Table 2.[22]

In column 8 of Table 2 we run the same regression as in Column 7 but for the IV sub-sample of 181 districts. The results are qualitatively similar to those in column 7. Finally, in column 9 of Table 2, we account for the endogeneity of both the explanatory variables and instrument $Brit_{i,j}$ and $Lit.Rate_{i,j}$ using their respective instruments ($Lapse_{i,j}$ and $Cath_{i,j}$ respectively). Panel B reports the first stage for $Brit_{i,j}$ and shows that both $Lapse_{i,j}$ and $Cath_{i,j}$ are significantly positively related with $Brit_{i,j}$. In panel C only $Cath_{i,j}$ is significantly positively related with $Lit.Rate_{i,j}$.[23] The coefficient on $Brit_{i,j}$ in the second stage is negative and significant and slightly larger in magnitude than the corresponding OLS coefficient in column 5. The literacy rate remains positively significant and is also larger in magnitude than the corresponding OLS coefficient in column 5. Therefore, even after instrumenting both direct British rule and literacy, we find that directly ruled districts have significantly less nightlights per capita in 1993.

---

[21] The Kleibergen and Paap (KP) F statistics for the first stage in column 7 is 15.73 which lies above the Stock and Yogo (2005) critical value for 15% maximal IV size of 8.96. In column 6, the KP F statistics is 7.72 which is above the Stock and Yogo (2005) critical value for 20% maximal IV size of 6.66.

[22] As mentioned in the Introduction, they use 2001 district boundaries and work with 500 districts while we use 1991 district boundaries and work with 412 districts which makes a direct comparison of results difficult.

[23] Since we instrument both the endogenous regressors in column 9 of Table 2, the Stock and Yogo (2005) critical values for column 9 are different from columns 6, 7, and 8 where we instrument only one endogenous regressor. The KP F statistics of 3.71 in the first stage is above the Stock and Yogo (2005) critical value for 25% maximal IV size of 3.63.

Note that, so far, our identification of the effect of direct British rule on overall development relies on both variations in districts across different modern Indian states and districts within modern Indian states. In the next set of regressions, we use state fixed effects to see if the results survive if we just rely on variations within modern Indian states. That is, in the exercises below we are comparing the levels of development of directly and indirectly ruled districts within the same modern state. Since the districts which have been a part of the same Indian state after independence have faced similar policies, it is possible that any effect of colonial policies will be diluted over time.

III.7. State Fixed effect regression results

The results of the estimation of equation (2) with state fixed effects are reported in Table 4. Relative to column 1 of Table 2, the coefficient on $Brit_{i,j}$ in column 1 of Table 4 is much smaller in magnitude suggesting that part of the variation in overall development across districts stems from the variation in state-level policies practiced since independence. That is, the difference in the level of development across directly and indirectly ruled districts is much less if they belong the same state than if they were in different states. However, on the inclusion of literacy rate in the subsequent columns of Table 4, we find that the coefficient on $Brit_{i,j}$ increases and is now much closer in magnitude to the respective coefficients in Table 2. Similar to the findings in Table 2, the increase in the magnitude of the coefficient on $Brit_{i,j}$ stems from the positive correlation between $Brit_{i,j}$ and $Lit.Rate_{i,j}$, and this correlation is much stronger within states than across states.

Columns 3 and 4 in Table 4 report the results of the neighborhood regressions. Restricting the neighborhood districts to be in the same modern Indian state restricts the sample to only 151 districts compared to 221 in Table 2. The results from the within state neighborhood regressions are qualitatively similar to those in columns 3 and 4 in Table 2. The other columns in Table 4 repeat the same regressions as in Table 2 but add the state fixed effects. In all the specifications except in column 9, the coefficient on $Brit_{i,j}$ is negative. The coefficient of $Brit_{i,j}$ in column 9, when we instrument both $Brit_{i,j}$ and $Lit.Rate_{i,j}$, is very large in size but fails to be statistically significant.[24]

[24] As a robustness check, we replicate our results in Tables 2 and 4 with an alternative measure of human capital, the share of higher education in each district in 1991. This is the measure of human capital used by Castello-Climent et al. (2018) in their study.

Overall, the results we obtain from Tables 2 and 4 confirm that the colonial institutions in the form of direct British rule are a major cause of disparity in present-day economic development. The results persist and become stronger after controlling for human capital suggesting that other mechanisms are at play. In the next section, we explore probable mechanisms, particularly differences in health and infrastructure levels, through which British direct rule affects present-day overall development.

III.8. Other Channels: Health and Physical infrastructure

If it is the case that the direct British rule led to underinvestment in health and physical infrastructure which has a persistent effect on economic development or directly ruled areas are still underinvesting in infrastructure, then the inclusion of these controls would reduce or eliminate the effect of direct British rule on nightlights per capita. Our measure of health is the infant mortality rate (IMR hereon) and we use road length per capita and railroad length per capita to capture physical infrastructure.[25]

The results from including these additional controls are reported in Table 5. The first 4 columns present estimates of regressions without state fixed effects while the last 4 columns include state fixed effects. In Column 1 we add IMR, roads per capita and railroad per capita to the regression reported in column 2 of Table 2. IMR is negatively associated with nightlights per capita but the coefficient is statistically insignificant. Both roads and railroads are positively associated with nightlights per capita and their coefficients are statistically significant. The coefficient of $Brit_{i,j}$ is slightly smaller in magnitude compared to column 2 in Table 2 but remains negative and significant. Column 2 of Table 5 produces the results form the sub-sample of neighboring districts and the results are similar to the corresponding result in column 5 of Table 2. Column 3 presents the OLS results for the IV sample and again they are similar to the corresponding results in column 5 of Table 2. Column 4 presents the IV results when $Brit_{i,j}$ is instrumented by *Lapse* and the IV results are similar to the corresponding IV results in column 6 of Table 2. Similarly, the results with state fixed effects in columns 5-8 are similar to the

The results are reported in Tables A1 and A2 in the online appendix and are very similar to those obtained in Tables 2 and 4 using the literacy rate.

[25] For information on data for IMR, road length per capita, and railroad length per capita data, refer to the data Appendix.

corresponding results in Table 4. Therefore, our main result that the directly ruled British districts have significantly lower nightlights per capita is robust to the inclusion of measures of health and physical infrastructure. The results also suggest that there are channels other than human capital, health, and physical infrastructure through which direct British rule has persistent effect on development.

In the next section we explore the role of the type of land tenure system adopted by the British in explaining the differences across directly and indirectly ruled districts.

III.9. Why are areas under direct rule doing worse?

Land revenue was the biggest source of government revenue during the colonial period (Iyer, 2010). The collection of land revenue was facilitated by land revenue collection systems referred to as the land tenure systems developed by the British. These land tenure systems essentially determined who had proprietary rights over the land. In the landlord-based system, the landlord paid a fixed amount of revenue to the British but was free to exploit the tenant. In the individual cultivator-based system, the cultivator was responsible for paying the revenue while in the village-based system the village was responsible for it. The landlord-based system has been shown to be more exploitative with persistent adverse effects on development. Looking at only the directly ruled areas, Banerjee and Iyer (2005) found that landlord districts had lower investments in agriculture and worse public good provision. Iyer (2010) also looked at indirectly ruled areas and found that there was not much difference in the public good provision between landlord and non-landlord districts in indirectly ruled areas. In her microlevel study of villages in 26 districts of UP, Pandey (2010) found that governance and educational outcomes were worse in villages in landlord districts compared to non-landlord districts.

In this section, we explore the implications of the land tenure system for overall development measured by nightlights per capita. We study three questions: 1) Do they affect the current level of development? 2) Do the effects differ across areas under direct and indirect rules? 3) Do the effects exist even after controlling for health, education, and physical infrastructure? To do this we use data from Iyer (2010) which has information on the proportion of land under non-landlord tenure system in each district. We denote the proportion of land under non-landlord-based tenure system in a district by *NLT* and augment the specification described in equation (1) with two additional explanatory variables, specifically $NLT_{i,j}$ and an interaction of $Brit_{i,j}$ and $NLT_{i,j}$, to obtain the following estimating equation.

$$y_{i,j} = \alpha_3 + \beta Brit_{i,j} + \theta NLT_{i,j} + \lambda\left(Brit_{i,j} \times NLT_{i,j}\right) + \gamma X_{i,j} + \varepsilon_{i,j} \dots (3)$$

Given the specification in equation (3), the omitted category is the indirectly ruled landlord districts. θ captures the difference in the level of development between indirectly ruled landlord and non-landlord districts. β captures the difference between directly ruled landlord districts and indirectly ruled landlord districts. $\beta + \theta + \lambda$ captures the difference between indirectly ruled landlord districts and directly ruled non-landlord districts.

In equation (3), the $NLT_{i,j}$ variable could potentially suffer from similar endogeneity issues that we discussed previously in the case of $Brit_{i,j}$. To tackle this selection bias, Banerjee and Iyer (2005) extend their analysis to an instrumental variable strategy wherein they use a dummy variable, that takes the value of 1 if the respective area was annexed between 1820 and 1856, as an instrument for the proportion of land area that had the non-landlord-based system. They argue that areas where the British took control of the land revenue collection between 1820 and 1856 are more likely to have majority of their land area fall under the non-landlord-based system. They find their OLS estimates to be biased downwards and treat their OLS results as the benchmark estimates in their paper due to the potential upward bias in their IV estimates. Based on the analysis by Banerjee and Iyer (2005), we maintain that endogeneity of land tenure institution is not a concern.

The data on land tenure systems in place is only available for a total of 356 modern 1991 districts. Out of 356 districts, 245 districts came under the direct British rule and 111 districts were part of native states and hence came under the indirect British rule. Of the 245 directly ruled districts, about 62% had a majority of their land area under the non-landlord-based system. For the 111 indirectly ruled districts it was about 64%. This suggests that, on average, the distribution of the type of land tenure system in place was largely similar across areas under direct and indirect rule. Figure 4 showcases the distribution of land tenure systems across directly and indirectly ruled districts with different land tenure systems.

Table 6 reports the results of estimating equation (3). Since the land tenure system data is only available for 356 districts, we replicate our benchmark specifications from column 1 of Table 2 with this restricted sample of 356 districts and report the results in columns 1 in Table 6. In column 2, we replace $Brit_{i,j}$ with $NLT_{i,j}$, the proportion of land under non-landlord-based tenure system, as the main explanatory variable. The coefficient of $NLT_{i,j}$ is negative and significant. The coefficient implies that a district that was completely under the landlord-based tenure system had 63% less nightlights per capita in 1993 relative to modern districts that were completely under the non-landlord-based tenure system. That is, landlord districts are doing much worse and other regressions show that this result is primarily driven by the landlord districts under direct rule. In column 3, we report the results obtained for the specification described in equation (3). In column 4 we include literacy rate as an additional control and in column 5 we add IMR and physical infrastructure measures as additional controls. The coefficient of $NLT_{i,j}$ is statistically insignificant in columns 3-5 even though it is quantitatively significant in column 3. That is, there is no robust evidence that the land tenure system has a significant effect on nightlights per capita in indirectly ruled districts. The coefficient of $Brit_{i,j}$ is negative and significant and the coefficient of interaction of $Brit_{i,j}$ and $NLT_{i,j}$ is positive and significant in columns 3-5. Based on the estimates reported in column 5, we can say the following. There is no significant difference between the non-landlord districts and landlord districts under indirect rule (estimate of θ is .048 but insignificant). Landlord districts under direct rule had 61.4% less nightlights per capita compared to landlord districts under indirect rule (estimate of β is -.95); Non-landlord districts under direct rule had 11.5% less nightlights per capita than the indirectly ruled landlord districts (estimate of $\beta + \theta + \lambda$ is -.21). Therefore, these results suggest that a large part of the difference in overall development between directly and indirectly ruled districts is driven by the poor performance of directly ruled landlord districts. That is, the landlord-based revenue system in combination with direct British rule has the most long-lasting adverse effect on development. [26]

As mentioned earlier, Banerjee and Iyer (2005) and Iyer (2010) found significant differences in the public good provision between landlord and non-landlord districts in directly ruled areas. Our results using nightlights data suggest differences in the overall development between landlord and non-landlord districts under direct rule. Moreover, the differences persist even after controlling for educational attainment, health, and physical infrastructure.

[26] We also try an alternative specification where *NLT* is an indicator variable which takes the value of 1 if district *i* in state *j* was historically part of an area where majority of the land area had non-landlord-based revenue collection system and a value of 0 otherwise. The results are qualitatively similar and are reported in the online appendix Table A5.

To check the robustness of above results we also did a neighborhood district analysis where we restricted our sample to only those districts which only had a neighboring district with a different land tenure system. To do the neighboring district analysis we had to use the binary classification of landlord/non-landlord districts as discussed in footnote 26. Figure 5 shows the distribution of lights per capita in adjacent landlord and non-landlord districts. Our sample size is reduced from 356 to 170 districts. The regression results are presented in columns 6-8 in Table 6. The key difference compared to the results in columns 3-5 is that the coefficient of the interaction term, $Brit_{i,j} \times NLT_{i,j}$, becomes smaller in magnitude and loses statistical significance. That is, the gap between landlord and non-landlord districts under direct rule becomes much smaller. However, it is still the case that the directly ruled landlord districts have significantly less nightlights per capita than the other 3 types of districts. We also tried to include state fixed effects in our regressions with land tenure, however, the results are less robust. This is mainly because the land tenure system varied more across districts which are parts of different modern Indian states than districts within modern Indian states.

As mentioned in the introduction, Banerjee and Iyer (2005) explain the adverse effects of the landlord-based system arising due to the concentration of power in the hands of an elite which fostered class-divide and conflict resulting in limited collective action. Pandey (2010) corroborates this hypothesis using microlevel evidence and suggests that the elite have “little stake” in the provision of public goods which mainly benefit the non-elite. Our findings that the landlord districts do much worse despite controlling for education, health, and infrastructure suggests that more work needs to be done in understanding the channels through which these colonial institutions are still affecting development. One possibility, considering the results of Pandey (2010) regarding the quality of education, is that our human capital variables (literacy rate or secondary schooling attainment of population) may not be capturing the quality of education. It is difficult to obtain district level data on the quality of education, however, there is one dataset which has some information on test scores of children in some districts of India which we make use of. The test score data are obtained from the India Human Development Survey 2005 (IHDS-I)[27]. As a part of this survey, children aged 8-11 in each household in the sample completed short reading,

[27] The IHDS consists of data from 41,554 households in 1,504 villages and 970 urban neighborhoods across India in 2005. The IHDS covers a comprehensive list of topics such as health, education, employment, economic status, marriage, fertility, gender relations, and social capital. It is a project carried out jointly by researchers from University of Maryland and the National Council of Applied Economic Research (NCAER), New Delhi.

writing and arithmetic tests that were developed by the IHDS team in collaboration with researchers from Pratham, India.[28] The tests were given in reading, math, and writing. We take the average of the scores across the three tests for each student to obtain a proxy for the quality of education in each district. They vary from 0.2 to 1 with a mean of 0.63 and a standard deviation of 0.14. One limitation is that the test score data are available for only 280 districts which reduces our sample size considerably. In column 9 in Table 6 we use the average test score as the dependent variable and find that test scores are slightly higher in the directly ruled districts compared to indirectly ruled districts. They are also slightly higher in non-landlord districts, however, the coefficient of the variable proportion of land under non-landlord-based tenure systemfails to be statistically significant. Therefore, the results do not provide a clear-cut evidence that the quality of education is worse in landlord districts. When we include test scores as a measure of human capital instead of literacy rate in column 10, the results are qualitatively similar to those in column 5 where literacy rate is used as a measure of human capital.[29] That is, nightlights per capita are significantly lower in directly ruled landlord districts compared to the other 3 types of districts.

In the next section, we examine the robustness of our main results.

IV. Robustness Exercises

In this section, we perform a series of robustness checks to examine the sensitivity of our results. For this purpose, we treat the specification in columns 1 and 5 of Table 5 as our baseline results (OLS regressions with and without state fixed effects) . In addition to these specifications, we also check the robustness of the results with land tenure systems in column 5 of Table 6.

IV.1. Measurement issues in nightlights data

Overglow and blooming of nightlights: A potential concern in the measurement of nightlights data is the spillover of nightlights from urban cities in close proximity. As a robustness check, we follow Chanda and Kabiraj (2020) and account for the distance to the closest major city from each district

---

[28] These tests were pretested to ensure comparability across languages.
[29] Since the IHDS test score data are available for 2005, the nightlights in the regression in column 10 are for year 2005.

to account for the overglow and blooming effect prevalent in measuring nightlights.[30] The results without state fixed effect are reported in column 1 and with state fixed effect in column 2 of Table 7. The coefficients are similar in magnitude and significance to those reported in columns 1 and 5 of Table 5. For the land tenure regressions, the results are reported in Table 8. Column 1 in Table 8 reports the results of adding this additional regressor to the regression reported in column 5 of Table 6. Again, the results are very similar.

Top-coding and Bottom censoring in Nightlights Data: The nightlights data (stable lights product) that we utilize in our main specification is censored at a DN of 63 at the top (top-coded) and 0 at the bottom. This implies that all the brightest areas have a DN of 63 resulting in negligible variation in their nightlight intensity measurement. In a similar manner, if nightlight intensity from the least bright areas does not meet the minimum requirement to be captured by the satellite, a DN of 0 is assigned to these areas.[31] The top-coding of nightlights data may limit variation in our outcome variable of $Nightlights\ Per\ Capita$ and could potentially affect the significance of our results. As a robustness check, we follow Chanda and Kabiraj (2020) and use an alternative nightlights dataset by Bluhm and Krause (2018) (BK hereon) that corrects for top-coding.[32] We obtain the corrected nighttime lights image for the year 1993 from BK and obtain the *Nightlights Per Capita* statistic for all the 466 districts in India as per 1991 Census.[33] Columns 3 and 4 in Table 7 report results from using this alternative measure of nightlights. Note that the results in columns 3 and 4 are very similar to the corresponding results in columns 1 and 2 in Table 7. Similarly, for the land tenure regressions, the results using nightlights from BK reported in column 2 of Table 8 are very similar to those in column 5 of Table 6. Therefore, our main results are robust to the use of an alternative nightlights data free of top-coding.

---

[30] We find the distance from the centroid of each district to the closest major city in ArcGIS using the 1993 district-level shapefile. The major cities as per the 1991 Census are Mumbai, Delhi, Bangalore, Kolkata, Chennai, Ahmedabad, Hyderabad, Pune, Surat, Kanpur, Jaipur, and Lucknow.

[31] See Henderson et al. (2012), Pinkovskiy and Sala-i-Martin (2016), Prakash, Rockmore, and Uppal (2019), and Chanda and Kabiraj (2020) for more information on the top-coding issue of nightlights data.

[32] The *Global Radiance Calibrated Nighttime Lights* product that is made publicly available by the NOAA's NGDC is also free of the top-coding issue. This product however is only available of years 1996, 1999, 2000, 2002, 2004, 2005, and 2010. Since we are interested in an outcome variable closer to the 1991 census year, we use the Bluhm and Krause (2018) nightlights dataset that is corrected for top-coding for all years spanning 1992-2013. The correlation between the *sum of lights* statistic obtained for all the 466 districts of India from both these datasets is about 0.987 for the year 2010.

[33] The Bluhm and Krause (2018) nightlights data was retrieved on Aug 12, 2020 from https://lightinequality.com/top-lights.

IV.2. Year-specific results

One limitation of our analysis is its cross-sectional nature and focus on the outcome in a particular year (1993). It is possible that results may vary for outcome variables from different years. As a robustness check, we use the natural log of nightlights (BK) per capita in 2013 as the outcome variable instead.[34] Columns 5 and 6 in Table 7 report the results obtained using per capita lights (BK) in year 2013. In both these columns, the coefficient on $Brit_{i,j}$ remains negative and significant. An interesting point to note is that the magnitude of the coefficient on $Brit_{i,j}$ decreases by 23.52% from column 3 to 5 and by 28.30% from column 4 to 6 as we change the year of the outcome variable from 1993 to 2013. This decrease in the magnitude of the coefficient indicates that the disparity in economic development across directly and indirectly ruled districts has decreased overtime since 1993. This is consistent with the evidence of convergence that we reported in Table 3 earlier. We find a similar pattern when we use the nightlights (BK) for 2013 in the regressions with land tenure system. Comparing the results in column 3 of Table 8 with the corresponding results in column 2 we find a substantial decline in the coefficients of $Brit_{i,j}$ and $Brit_{i,j} \times NLT_{i,j}$. While the coefficient of $Brit_{i,j}$ in column 2 implied that in 1993 the landlord districts under direct British rule had 60% less lights per capita than the landlord districts under indirect rule, by year 2013 this was down to 32%.

We validate these results further by using nightlights data of a higher resolution (15 arc-second, about 0.5 km) captured by the JPSS-VIIRS (Elvidge et al., 2021) and was recently made available.[35] This data spans 2012-2020 and hence is available for more recent years. Compared to the DMSP-OLS nightlights data, the JPSS-VIIRS nightlights data is free of over-saturation and has the provision for onboard calibration which significantly enhances the data quality. For the purpose of our analysis, we replicate specifications in columns 5 and 6 of Table 7 using nightlights (JPSS-VIIRS) per capita[36] in 2013, 2016, and 2019 as the dependent variable and report results in

[34] Considering the history of district bifurcations and boundary changes, we refer to Kumar and Somanathan (2009) and Law (2016) in matching the 2011 district boundaries to the 1991 district boundaries and obtain the population of all the 466 districts (as per the 1991 Census) in 2011. Since population data in 2013 at the district level is not available, we use 2011 population data and the *Sum of lights* statistic for 2013 to construct the *Nightlights Per Capita* variable for year 2013 for all the 466 districts.

[35] The JPSS-VIIRS nightlights dataset has 8 data products for each year. In our analysis, we use the masked average radiance product in which background, biomass burning, and aurora have been zeroed out. The JPSS-VIIRS nightlights data was retrieved on April 8, 2021 from https://eogdata.mines.edu/nighttime_light/annual/v20/. Similar results are obtained using their unmasked average radiance data.

[36] Due to unavailability of district-level population data for years 2013, 2016, and 2019, we use the 2011 district-level census population to compute nightlights per capita for these years in Table A6.

Table A6 in the online appendix. In all the 6 columns in Table A6, the coefficient on $Brit_{i,j}$ remains negative and significant. We also observe a decrease in size of the coefficient on $Brit_{i,j}$ from 2013-2019 indicating a decrease in disparity in economic development across directly and indirectly ruled districts overtime since 2013. Since the quality of the JPSS-VIIRS nightlights data is very different from the nightlights data for earlier years, we do not extend our growth analysis to later years.

IV.3. Sample Size

In our main analysis, we follow the common practice of focusing on 412 districts spread across 18 major states of India and leave out 54 districts from the remaining 14 states of India for reasons including quality of current data, problems in matching current and historic district boundaries, and availability of data. In this section, we extend our main analysis and include these excluded districts in our sample. Of the remaining 54 districts, 9 districts were historically colonies of the Portuguese and the French and hence we exclude them from our sample resulting in an extended total sample size of 457 districts.[37] From our sample of 457 districts, 295 districts were under direct British rule while 162 were under indirect British rule. We replicate our specifications in columns 1 and 2 of Table 7 using this sample of 457 districts and report the corresponding results in columns 7 and 8 of Table 7. The results remain robust in the augmented sample.

IV.4. Duration of colonial rule

Feyrer and Sacerdote (2009) find that the variation in the amount of time spent as a colony has a differential impact on present-day economic development. Considering this, we examine the robustness of our results to an alternate measure of direct British rule. We resort to using the number of years spent under direct British rule as the main explanatory variable in place of the $Brit_{i,j}$ indicator variable. We obtain data on the duration of direct British rule from Iyer (2010). We replicate specifications in columns 1 and 2 of Table 7 with $DuraBrit_{i,j}$, which represents the number of years a district i in state j was under direct British rule, in place of $Brit_{i,j}$ as the main

[37] For more information on the institutional data for the remaining 45 districts, refer to the data appendix.

explanatory variable and report the results in columns 9 and 10 of Table 7. The coefficient on $DuraBrit_{i,j}$ is negative and significant in both these columns suggesting that the duration of direct British rule has a negative effect on nightlights per capita in 1993. In a similar manner, we do this exercise for the specification where land tenure system is included and reported in Table 8. The last column in Table 8 replaces $Brit_{i,j}$ with $DuraBrit_{i,j}$ and $Brit_{i,j} \times NLT_{i,j}$ with $DuraBrit_{i,j} \times NLT_{i,j}$ as the main explanatory variables and the results are similar to those in column 1.

IV.5. Additional Controls

Next, we check if our main results are robust to the inclusion of additional controls, specifically, terrain ruggedness, historical controls, and agricultural suitability index.[38] Considering the blooming and overglow effect prevalent in nightlights data (Chanda and Kabiraj, 2020) as well as the fact that measurement of nightlights data is affected by the terrain type of a region (Pinkovskiy and Sala-i-Martin, 2016), we include a measure of terrain ruggedness as an additional control in our specifications to account for these issues prevalent in the measurement of nightlights. Next, we include a measure of agricultural land suitability index in our set of controls to account for exogenous differences in agricultural suitability across districts as well as the selectivity of the British towards areas more favorable for agriculture. In addition to these two controls, we also include a set of historical controls to capture any possible differences in the initial conditions across modern districts that may bias our main results.

We augment the specifications in columns 1 and 2 of Table 7 with these controls and report the resulting estimates in Table A7 in the online appendix. We also include these controls in the specification of column 1 in Table 8 and report the corresponding results in Table A8 in the online appendix. Our main results are robust to the inclusion of these additional controls.

IV.6. Alternate proxy of development

[38]The details are provided in the data appendix.

In this section, we use an alternate proxy of development, specifically the average consumption per capita in each district, instead of nightlights per capita and examine if our main results still hold. We obtain the 1993-1994 district-level measure of average consumption per capita from Topalova (2010). One limitation of this data is that while it is available at the district-level for the rural sector, for the urban sector it is only available at the regional level. Since our main analysis is at the district-level, we perform this robustness test only for the rural sector. We match districts (1991 level) in our sample to the districts (1987 level) in Topalova (2010) and follow the method of Castello-Climent et al. (2018) in assigning data to the unmatched districts.[39] We replicate the specifications in columns 1 and 2 of Table 7 using log of consumption per capita (rural district) as the outcome variable and report the resulting estimates in Table A9 in the online appendix. Column 3 in Table A9 uses log of per capita consumption as the dependent variable in the land tenure regression and the results are similar to those obtained using nightlights. Overall, the results are robust to using this alternative proxy of overall economic development.

V. Concluding Remarks

This paper has studied the implications of two historical institutions, direct British rule, and the heterogeneous land tenure institutions implemented by the British, on disparity in present day development using district level data from India. While these institutions have been studied previously by Banerjee and Iyer (2005) and Iyer (2010), due to lack of data they were unable to look at the implications of these institutions on a comprehensive measure of development. Following the recent literature, we have used nightlights per capita as a proxy for district level per capita income. We find that modern districts that were historically under direct British rule had 39.47% less nightlights per capita in 1993 relative to modern districts that were historically under indirect British rule. Though the disparity in economic development across directly and indirectly ruled districts has decreased overtime since 1993 till 2019, the gap is still significantly large (18-21%) and persists even after including other controls such as educational attainment, health, and physical infrastructure. Looking at the growth pattern during 1993 to 2013, a period of rapid

[39] Unmatched districts are a result of boundary changes between 1987 and 1991. In matching 1931 level districts to 2001 level districts, in the case of unmatched districts or districts with missing data, Castello-Climent et al. (2018) assign the same data value to districts that were historically part of the same geographic boundary. We follow this method and assign the same data value to districts that were part of the same geographic boundary in 1987.

growth following the liberalization that began in 1991, we find that areas that were under direct British rule had a 1.84% lower annual growth rate compared to indirectly ruled areas. We also find that areas under direct British rule were converging at a rate of 2% per year while areas that were under indirect British rule were converging at a rate of 5.7% per year.

Looking at the impact of land tenure institutions, we find that much of the development gap between areas under indirect rule and direct rule can be accounted for by the adverse effect of landlord-based revenue collection system in the directly ruled areas. The development gap remains even after controlling for education, health, and physical infrastructure. The results indicate that these institutions have persistent effects on development through subtle channels that are not fully captured by variables like education, health, and physical infrastructure. Future research should attempt to uncover these subtle channels.

References

Acemoglu, Daron, Johnson, Simon, and Robinson, James A. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review,* 91 (5): 1369-1401.

Acemoglu, Daron, Gallego, Francisco A., and Robinson, James A. 2014. "Institutions, Human Capital, and Development." *Annual Review of Economics,* 6: 875-912.

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

Banerjee, Abhijit, Iyer, Lakshmi, and Somanathan, Rohini. 2005. "History, Social Divisions, and Public Goods in Rural India." *Journal of the European Economic Association,* 3 (2-3): 639–647.

Barro, R., and Sala-i-Martin, X. 1992. "Convergence." *Journal of Political Economy,* 100 (2): 223–251.

Becker SO, and Woessmann L. 2009. "Was Weber wrong? A human capital theory of Protestant economic history." *The Quarterly Journal of Economics,* 124 (2): 531–596.

Besley, Timothy and Burgess, Robin. 2000. "Land Reform, Poverty Reduction, and Growth: Evidence from India." *The Quarterly Journal of Economics,* 115 (2): 389–430.

Bluhm, Richard and Krause, Melanie. 2018. "Top Lights-Bright Spots and Their Contribution to Economic Development." *CESifo Working Paper No. 7411.*

Castelló-Climent, Amparo, Chaudhary, Latika, and Mukhopadhyay, Abhiroop. 2018. "Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India." *The Economic Journal,* 128 (616): 3039–3075.

Chanda, A., & Kabiraj, S. 2020. "Shedding light on regional growth and convergence in India." *World Development,* 133: 104961.

Chanda, Areendam, Cook, C. Justin, and Putterman, Louis. 2014. "Persistence of Fortune: Accounting for Population Movements, There Was No Post-Columbian Reversal." *American Economic Journal: Macroeconomics,* 6 (3): 1-28.

Chen, Xi, and Nordhaus, William D.. 2011. "Using luminosity data as a proxy for economic statistics." *Procedings of the National Academy of Sciences of the United States of America,* 108 (21) 8589-8594.

Dell, M. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica,* 78: 1863-1903.

Dell, Melissa and Olken, Benjamin A. 2019. "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java." *The Review of Economic Studies,* 87 (1): 164–203.

Donaldson, Dave, and Storeygard, Adam. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives,* 30 (4): 171-98.

Elvidge, C.D, Zhizhin, M., Ghosh, T., Hsu, FC., and Taneja, J. 2021. "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019." *Remote Sensing*, 13 (5): 922.

Feyrer, James, and Sacerdote, Bruce. 2009. "Colonialism and Modern Income: Islands as Natural Experiments." *The Review of Economics and Statistics,* 91. 245-262.

Gallego FA, and Woodberry RD. 2009. "Christian missionaries and education in former colonies: how institutions mattered." *Working Paper 339, Inst. Econ., Pontif. Univ. Catol. Chile, Santiago.*

Gallego FA, and Woodberry RD. 2010. "Christian missionaries and education in former African colonies: how competition mattered." *Journal of African Economies,* 19: 294–329.

Glaeser, E., La Porta, R., Lopez-De-Silanes, F., and Shleifer, A. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth,* 9 (3): 271-303.

Government of India. 1997. "District level estimates of fertility and child mortality for 1991 and their inter relations with other variables". *Occasional Paper No. 1 of 1997, Registrar General, India.*

Henderson, J. Vernon, Storeygard, Adam, and Weil, David N. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review,* 102 (2): 994-1028.

India State-Level Disease Burden Initiative Child Mortality Collaborators. 2020. "Subnational mapping of under-5 and neonatal mortality trends in India: the Global Burden of Disease Study 2000-17." *Lancet,* 395 (10237): 1640-1658.

Iyer, L. 2010. "Direct versus Indirect Colonial Rule in India: Long-Term Consequences." *The Review of Economics and Statistics,* 92 (4): 693-713.

Kumar, H. and Somanathan, R. 2009. "Mapping Indian districts across census years, 1971-2001." *Economic and Political Weekly,* 44 (41/42): 69-73.

Law, Gwillim. 2016. "Districts of India". January 29. http://www.statoids.com/yin.html

Lipset SM. 1959. "Some social requisites for democracy: economic development and political legitimacy." *The American Political Science Review,* 53 (1): 69–105.

Nunn, Nathan. 2008. "The Long-term Effects of Africa's Slave Trades." *The Quarterly Journal of Economics,* 123 (1): 139–176.

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

Pandey, Priyanka. 2010. "Service Delivery and Corruption in Public Services: How Does History Matter?" *American Economic Journal: Applied Economics,* 2 (3): 190-204.

Pinkovskiy, Maxim, and Sala-i-Martin, Xavier. 2016. "Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate." *The Quarterly Journal of Economics,* 131 (2): 579–631.

Prakash, Nishith, Rockmore, Marc, and Uppal, Yogesh. 2019. "Do criminally accused politicians affect economic outcomes? Evidence from India." *Journal of Development Economics,* 141: 102370.

Ramankutty, N., J.A. Foley, J. Norman, and K. McSweeney. 2001. "The global distribution of cultivable lands: current patterns and sensitivity to possible climate change." *Submitted to Global Ecology and Biogeography.*

Rodrik, D., Subramanian, A., and Trebbi, F. 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development." *Journal of Economic Growth,* 9, 131–165.

Shah, Manisha, and Steinberg, Bryce. (2017). Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital. Journal of Political Economy. 125 (2). 000-000. 10.1086/690828.

Stock, J.H., and Yogo, M. 2005. *Testing for weak instruments in linear IV regression*. In: Stock, J.H., Andrews, D.W.K. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg, Chapter 5. Cambridge University Press.

Topalova, Petia. 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *American Economic Journal: Applied Economics,* 2 (4): 1-41.

Vanneman, Reeve, and Barnes, Douglas. 2000. "Indian District Data, 1961-1991: machine-readable data file and codebook." Center on Population, Gender, and Social Inequality, College Park, Maryland. http://vanneman.umd.edu/districts/index.html

Verghese, Ajay. 2019. "Colonialism, Landlords, and Public Goods Provision in India: A Controlled Comparative Analysis." *The Journal of Development Studies,* 55 (7): 1345-1363.

Willmott, C. J., and Matsuura, K. 2001. "Terrestrial Air Temperature and Precipitation: Gridded Monthly Time Series (1900 - 2017)". University of Delaware. http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html.

Woodberry RD. 2004. "The shadow of empire: Christian missions, colonial policy, and democracy in postcolonial societies." *PhD Thesis, Dep. Sociol., Univ. N. C., Chapel Hill.*

Woodberry RD. 2011. *Religion and the spread of human capital and political institutions: Christian missions as a quasi-natural experiment.* In: The Oxford Handbook of the Economics of Religion. New York: Oxford University Press. ed. R McCleary: 111–31.

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Log Lights Per Capita 1993	412	-4.955861	1.258763	-13.05183	-2.771338
Log Lights Per Capita 1993 BK	412	-4.941857	1.260172	-13.05183	-2.768921
Log Lights Per Capita 2013 BK	412	-4.282473	0.9082515	-6.680486	-2.393828
Growth in Lights Per Capita 1993-2013	412	0.0316757	0.042288	-0.0562707	0.3507649
Log Lights Density 1993	412	0.7410932	1.575447	-9.549381	4.401308
Log Lights Per Capita 2005	412	-4.74033	1.057755	-7.727509	-2.913558
Log Lights Per Capita 2013 VIIRS (masked)	412	-5.408443	0.9333231	-8.444502	-3.446018
Log Lights Per Capita 2016 VIIRS (masked)	412	-5.169811	0.7263191	-7.949014	-3.276721
Log Lights Per Capita 2019 VIIRS (masked)	412	-4.791126	0.6047089	-7.002578	-2.898178
Brit (direct rule dummy)	412	0.6432039	0.4796363	0	1
Lapse (direct rule instrument dummy)	181	0.1049724	0.3073681	0	1
Non-landlord proportion	356	0.5437063	0.4331245	0	1
Brit x Non-landlord proportion	356	0.3570715	0.4308896	0	1
Non-landlord dummy	356	0.5533708	0.4978431	0	1
Brit x Non-landlord dummy	356	0.3539326	0.4788616	0	1
Duration under direct rule (x 0.01)	412	0.9601497	0.7019899	0	2.57
Literacy Rate 1991	412	0.5017138	0.1478411	0.1900832	0.957199
Share Higher Education 1991	412	0.1307432	0.0640106	0.0339682	0.4342449
Share Primary & Middle Education 1991	412	0.1997638	0.0878713	0.054044	0.5639171
Infant Mortality Rate (IMR) 1991	412	78.44927	28.63653	22	166
Log Road Length (km) Per Capita 1992	412	-7.549692	0.8039137	-11.796	-3.259715
Log Railroad Length (km) Per Capita 1992	412	-9.095702	1.360772	-14.28573	-6.351312
Avg. Test Score 2005	280	0.6332968	0.1426919	0.2034314	1
Cath Miss (dummy for catholic missionary)	412	0.3470874	0.4766226	0	1
Area (sq. km) 1991	412	7207.481	5838.631	174	45652
Log Population 1991	412	14.30509	0.7298461	10.35118	16.11066
Avg. Temp (celsius) 1900-1993	412	24.5877	4.375674	-4.630096	28.79043
Avg. Precipitation (mm) 1900-1993	412	1246.673	722.7811	174.5461	4173.867
Avg. Elevation (m)	412	429.7069	692.1129	4.139117	4912.925
Log River Length (km) Per Capita 1992	412	-7.377772	0.9946077	-16.11066	-2.689159
Latitude	412	23.28523	5.991527	8.30512	34.53142
Min. Dist to Closest Major City (100 km)	412	2.642268	1.574567	0	8.844524
Dist. to Nearest Coast (100 km)	412	4.391992	3.279918	0.044902	13.28298
Terrain Ruggedness (100 m)	412	0.768727	1.421245	0.0282508	8.527494
Agricultural Suitability Index	377	0.5542667	0.2297723	0.0026452	0.972
Share of Urban Population 1931	394	0.1096216	0.0746501	0	0.4952972
Share of Brahman Population 1931	394	0.055564	0.0418288	0.0015112	0.2703725
Share of Tribal Poplation 1931	394	0.0322377	0.0807637	0	0.689738
Railway (dummy for presence of railway) 1909	394	0.8020305	0.3989757	0	1
Population Density 1931	394	309.3752	235.2362	33.88353	2105.109
European Population 1931	394	1427.882	2760.937	0	17699
Log Mean Consumption Per Capita (Rural Sector) 1993	396	5.63265	0.2500862	5.028798	6.498956
Log Mean Consumption Per Capita (Urban Sector) 1993	398	6.053725	0.1890757	5.478734	6.673213

Table 2: Institutions, Human Capital, Long-Run Development - OLS and IV Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS and Second-stage Regressions									
	Dependent variable: log lights per capita in 1993								
	Full Sample		Neighboring Districts		IV Sample		Full Sample	IV Sample	
	OLS	OLS	OLS	OLS	OLS	2SLS	2SLS	2SLS	2SLS
Brit direct rule	-0.502**	-0.584***	-0.164	-0.361***	-0.625***	-0.629**	-0.646***	-0.650***	-0.671**
	(0.204)	(0.158)	(0.127)	(0.123)	(0.146)	(0.267)	(0.157)	(0.148)	(0.307)
Literacy rate 1991		4.314***		3.118***	3.205***	3.206***	7.581***	4.112*	4.147*
		(0.790)		(0.535)	(0.633)	(0.616)	(1.714)	(2.290)	(2.439)
R-squared	0.372	0.538	0.479	0.587	0.580	0.580	0.443	0.573	0.573
Panel B: First-stage Regressions for Brit direct rule									
	Dependent variable: Brit direct rule indicator								
Lapse						0.595***			0.594***
						(0.184)			(0.189)
Literacy rate 1991						0.150			
						(0.434)			
Catholic missionary									0.101**
									(0.050)
R-squared						0.394			0.401
Panel C: First-stage Regressions for Literacy rate									
	Dependent variable: Literacy rate in 1991								
Catholic missionary							0.059***	0.070***	0.071***
							(0.015)	(0.025)	(0.024)
Brit direct rule							0.008	0.018	
							(0.027)	(0.051)	
Lapse									0.023
									(0.043)
R-squared							0.378	0.431	0.431
Panel D: IV statistics									
K-P LM stat (p-value) (Underidentif. test)						0.1356	0.0046	0.0085	0.0071
K-P F stat (Weak Identif. test)						10.399	15.732	7.725	3.715
Observations	412	412	221	221	181	181	412	181	181
Brit Instrumented	NO	NO	NO	NO	NO	YES	NO	NO	YES
Lit. rate Instrumented	NO	NO	NO	NO	NO	NO	YES	YES	YES

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010). We include geographical controls in all the columns. Geographical controls include area in 1991, average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast.

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Effect of Direct Rule on Economic Growth from 1993-2013

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: growth in nightlights per capita from 1993-2013						
	Full Sample			Districts under Direct Rule		Districts under Indirect Rule	
	OLS	OLS	OLS	OLS	OLS	OLS	OLS
log lights per capita in 1993	-0.0231***	-0.0243***	-0.0261***	-0.0165***	-0.0206***	-0.0341***	-0.0341***
	(0.0027)	(0.0044)	(0.00416)	(0.0022)	(0.00239)	(0.0027)	(0.00272)
Brit direct rule		-0.0184**	-0.0210***				
		(0.0075)	(0.00737)				
Literacy rate 1991			0.0543**		0.0749***		0.000146
			(0.0241)		(0.0178)		(0.0132)
R-squared	0.471	0.513	0.546	0.314	0.392	0.764	0.764
Observations	412	412	412	265	265	147	147

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010) only for columns 2 and 3.
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Institutions, Human Capital, Long-Run Development - OLS and IV Estimates with State Fixed-Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS and Second-stage Regressions									
	Dependent variable: log lights per capita in 1993								
	Full Sample		Neighboring Districts		IV Sample		Full Sample	IV Sample	
	OLS	OLS	OLS	OLS	OLS	2SLS	2SLS	2SLS	2SLS
Brit direct rule	-0.220***	-0.428***	-0.104	-0.238***	-0.632***	-0.689*	-0.633***	-0.576**	-0.618
	(0.0838)	(0.0995)	(0.0975)	(0.0844)	(0.238)	(0.379)	(0.164)	(0.294)	(0.489)
Literacy rate 1991		3.206***		1.820*	3.488***	3.526***	6.377***	2.559	2.641
		(0.704)		(0.949)	(1.200)	(1.144)	(1.987)	(2.594)	(2.824)
R-squared	0.628	0.673	0.759	0.776	0.646	0.646	0.629	0.642	0.643
Panel B: First-stage Regressions for Brit direct rule									
	Dependent variable: Brit direct rule indicator								
Lapse						0.526***			0.533***
						(0.182)			(0.192)
Literacy rate 1991						0.420			
						(0.349)			
Catholic missionary									0.077
									(0.058)
R-squared						0.660			0.657
Panel C: First-stage Regressions for Literacy rate									
	Dependent variable: Literacy rate in 1991								
Catholic missionary							0.045***	0.055***	0.058***
							(0.010)	(0.019)	(0.018)
Brit direct rule							0.054***	0.048	
							(0.014)	(0.033)	
Lapse									0.035
									(0.029)
R-squared							0.697	0.739	0.735
Panel D: IV statistics									
K-P LM stat (p-value) (Underidentif. test)						0.0579	0.0021	0.0099	0.0115
K-P F stat (Weak Identif. test)						9.393	20.591	9.474	4.339
Observations	412	412	151	151	181	181	412	181	181
Brit Instrumented	NO	NO	NO	NO	NO	YES	NO	NO	YES
Lit. rate Instrumented	NO	NO	NO	NO	NO	NO	YES	YES	YES

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010). We include geographical controls and state fixed-effects in all the columns. Geographical controls include area in 1991, average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast.

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Other Channels (Health and Infrastructure) - OLS and IV Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS and Second-stage Regressions								
	Dependent variable: log lights per capita in 1993							
	Full Sample	Neighboring Districts	IV Sample		Full Sample	Neighboring Districts	IV Sample	
	OLS	OLS	OLS	2SLS	OLS	OLS	OLS	2SLS
Brit direct rule	-0.524***	-0.345***	-0.507***	-0.708**	-0.341***	-0.242**	-0.480**	-0.806**
	(0.155)	(0.114)	(0.136)	(0.284)	(0.106)	(0.0934)	(0.186)	(0.377)
Literacy rate 1991	4.088***	2.633***	2.870***	2.999***	2.622***	1.645**	3.214***	3.534***
	(0.965)	(0.565)	(0.643)	(0.662)	(0.670)	(0.815)	(1.063)	(1.072)
IMR	-0.00316	-0.00618**	-0.00370	-0.00261	-0.00683**	0.000681	-0.00758*	-0.00539
	(0.00289)	(0.00239)	(0.00256)	(0.00270)	(0.00323)	(0.00361)	(0.00405)	(0.00344)
Log Road Length	0.399***	0.121	0.351***	0.321***	0.195*	-0.155	0.378**	0.363**
Per Capita	(0.114)	(0.155)	(0.127)	(0.115)	(0.106)	(0.158)	(0.173)	(0.154)
Log Railroad Length	0.0897**	0.0160	0.104**	0.103**	0.0862***	0.0971**	0.0308	0.0314
Per Capita	(0.0369)	(0.0475)	(0.0469)	(0.0449)	(0.0274)	(0.0376)	(0.0510)	(0.0476)
R-squared	0.566	0.608	0.603	0.601	0.689	0.786	0.663	0.659
Panel B: First-stage Regressions for Brit direct rule								
	Dependent variable: Brit direct rule indicator							
Lapse				0.571***				0.518***
				(0.180)				(0.161)
Literacy rate 1991				0.358				0.644*
				(0.441)				(0.378)
IMR				0.00358*				0.00588**
				(0.00191)				(0.00209)
Log Road Length				-0.214**				-0.107**
Per Capita				(0.104)				(0.046)
Log Railroad Length				0.0046				0.0068
Per Capita				(0.0202)				(0.0187)
R-squared				0.470				0.710
Panel C: IV statistics								
K-P LM stat (p-value) (Underidentif. test)				0.1064				0.0427
K-P F stat (Weak Identif. test)				10.060				11.644
Observations	412	221	181	181	412	151	181	181
Brit Instrumented	NO	NO	NO	YES	NO	NO	NO	YES
Lit. rate Instrumented	NO	NO	NO	NO	NO	NO	NO	NO
State FE	NO	NO	NO	NO	YES	YES	YES	YES

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010). We include geographical controls in all the columns. Geographical controls include area in 1991, average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of direct rule stemming from differences in land tenure institutions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Full Sample (Non-landlord proportion)					Neighboring landlord and non-landlord districts (Non-landlord dummy)			Quality of Human Capital (Non-landlord prop.)	
	Dependent variable: log lights per capita in 1993								Dep. Var: Avg. Test Score 2005	Dep. Var: log lights per capita 2005
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Brit direct rule	-0.513***		-1.043***	-1.065***	-0.950***	-0.517*	-0.676***	-0.616***	0.0685**	-0.665**
	(0.184)		(0.283)	(0.218)	(0.221)	(0.266)	(0.235)	(0.215)	(0.0316)	(0.294)
Non-landlord system		0.983***	0.215	-0.00810	0.0487	0.224	0.0733	0.0720	0.0545	0.234
		(0.308)	(0.154)	(0.135)	(0.141)	(0.141)	(0.137)	(0.147)	(0.0465)	(0.141)
Brit direct rule ×			1.007***	0.798***	0.689***	0.148	0.0509	0.0834	-0.00719	0.638*
Non-landlord system			(0.342)	(0.243)	(0.248)	(0.284)	(0.252)	(0.266)	(0.0470)	(0.356)
Literacy rate 1991				3.223***	3.238***		3.638***	3.649***		
				(0.591)	(0.725)		(0.435)	(0.670)		
IMR					-0.000941			0.000134		-0.00345
					(0.00346)			(0.00373)		(0.00329)
Log Road Length					0.264**			0.301***		0.411***
Per Capita					(0.101)			(0.103)		(0.103)
Log Railroad Length					0.0615			0.0906*		0.0856
Per Capita					(0.0372)			(0.0476)		(0.0522)
Avg. Test Score 2005										1.140***
										(0.345)
R-squared	0.301	0.391	0.456	0.560	0.575	0.305	0.506	0.547	0.181	0.493
Observations	356	356	356	356	356	170	170	170	248	248

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010). We include geographical controls in all the columns. Geographical controls include area in 1991, average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast.

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Robustness I - Effect of Direct British Rule (Measurement Issues, Year-Specific Results, Sample Size, Duration of Colonial Rule)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: log lights per capita in 1993		Dependent variable: log lights per capita in 1993bk		Dependent variable: log lights per capita in 2013bk		Dependent variable: log lights per capita in 1993		Dependent variable: log lights per capita in 1993	
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Brit direct rule	-0.503***	-0.308***	-0.511***	-0.318***	-0.390***	-0.228**	-0.562***	-0.287**		
	(0.136)	(0.0969)	(0.139)	(0.0957)	(0.120)	(0.0869)	(0.135)	(0.117)		
Duration of Brit direct rule (x 1/100)									-0.410***	-0.199***
									(0.107)	(0.0685)
Literacy rate 1991	3.864***	2.597***	3.957***	2.733***	3.483***	3.043***	3.512***	2.208***	3.556***	2.474***
	(0.802)	(0.643)	(0.807)	(0.641)	(0.518)	(0.407)	(0.873)	(0.797)	(0.682)	(0.636)
IMR	-0.00322	-0.00701**	-0.00293	-0.00698**	-0.00530**	-0.00486**	-0.00417	-0.00871*	-0.00377	-0.00742**
	(0.00284)	(0.00346)	(0.00287)	(0.00345)	(0.00213)	(0.00242)	(0.00312)	(0.00476)	(0.00266)	(0.00346)
Log Road Length Per Capita	0.466***	0.235**	0.437***	0.206*	0.518***	0.250***	0.405**	0.0714	0.405***	0.224*
	(0.115)	(0.114)	(0.119)	(0.112)	(0.0957)	(0.0930)	(0.173)	(0.180)	(0.114)	(0.116)
Log Railroad Length Per Capita	0.0946***	0.0884***	0.0954***	0.0885***	0.0421	0.0316	0.0508	0.0671	0.104***	0.0906***
	(0.0350)	(0.0282)	(0.0350)	(0.0277)	(0.0311)	(0.0266)	(0.0435)	(0.0419)	(0.0332)	(0.0284)
R-squared	0.598	0.695	0.601	0.700	0.598	0.732	0.545	0.662	0.608	0.693
Observations	412	412	412	412	412	412	457	457	412	412
Dist. to closest major city	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
State FE	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010). We include geographical controls in all the columns. Geographical controls include area in 1991, average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Robustness I - Direct British Rule and Land Tenure Institutions (Measurement Issues, Year-Specific Results, Duration of Colonial Rule)

	(1)	(2)	(3)	(4)
	Dependent variable: log lights per capita in 1993	Dependent variable: log lights per capita in 1993bk	Dependent variable: log lights per capita in 2013bk	Dependent variable: log lights per capita in 1993
	OLS	OLS	OLS	OLS
Brit direct rule	-0.897***	-0.919***	-0.387**	
	(0.197)	(0.203)	(0.189)	
Duration of Brit direct rule				-0.668***
				(0.138)
Non-landlord proportion	0.0333	0.0245	0.224	-0.0769
	(0.146)	(0.146)	(0.179)	(0.161)
Brit direct rule ×	0.655***	0.672***	0.0894	
Non-landlord proportion	(0.227)	(0.234)	(0.213)	
Dur. of direct rule ×				0.490***
Non-landlord proportion				(0.157)
Literacy rate 1991	3.222***	3.300***	3.253***	2.955***
	(0.669)	(0.677)	(0.537)	(0.604)
IMR	-0.000686	-0.000378	-0.00455*	-0.00204
	(0.00340)	(0.00346)	(0.00239)	(0.00318)
Log Road Length	0.339***	0.301**	0.465***	0.285**
Per Capita	(0.117)	(0.119)	(0.109)	(0.116)
Log Railroad Length	0.0839**	0.0841***	0.0189	0.0794**
Per Capita	(0.0319)	(0.0318)	(0.0306)	(0.0319)
R-squared	0.597	0.601	0.618	0.608
Observations	356	356	356	356
Dist. to closest major city	YES	YES	YES	YES

Note: Robust standard errors in parentheses. Standard errors are clustered at the level of the native state (Iyer, 2010). In columns 7 and 8, we use duration of direct British rule (x 1/100) as the main explanatory variable instead of the British dummy. We include geographical controls in all the columns. Geographical controls include area in 1991, average annual temperature (1900-1993), average annual rainfall (1900-1993), average elevation, latitude, log river length per capita, and distance to nearest coast.

*** p<0.01, ** p<0.05, * p<0.1

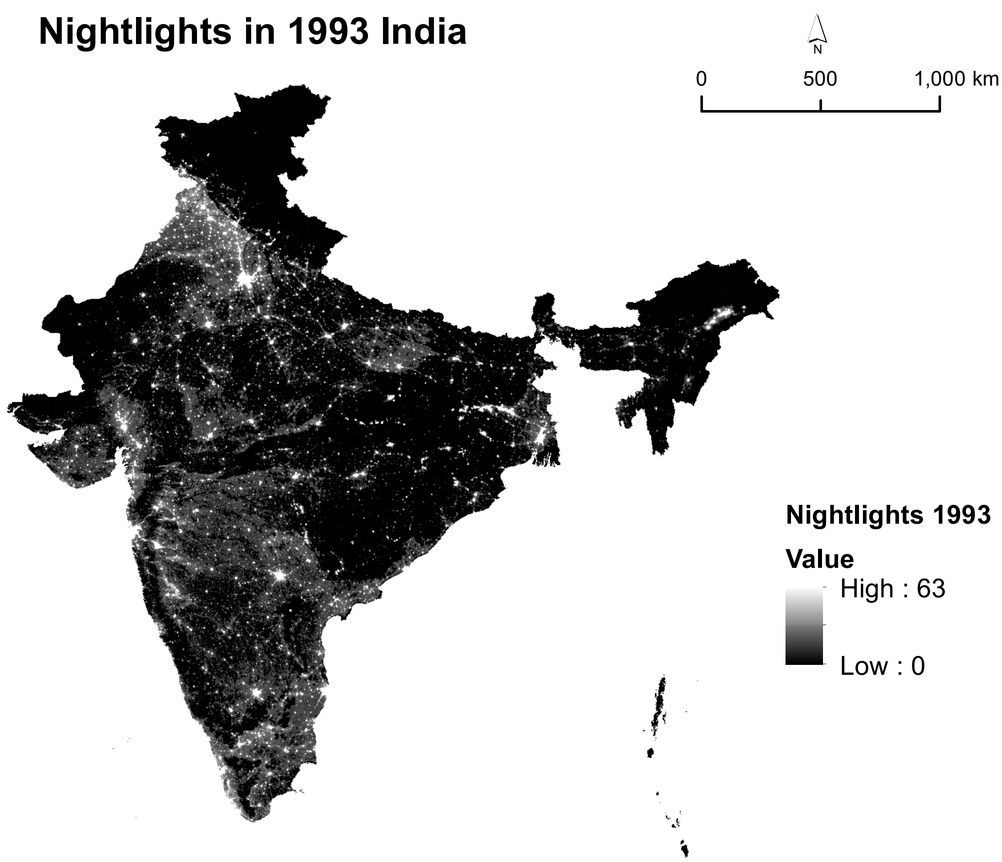

Figure 1: Raw Nightlights Image of India for year 1993

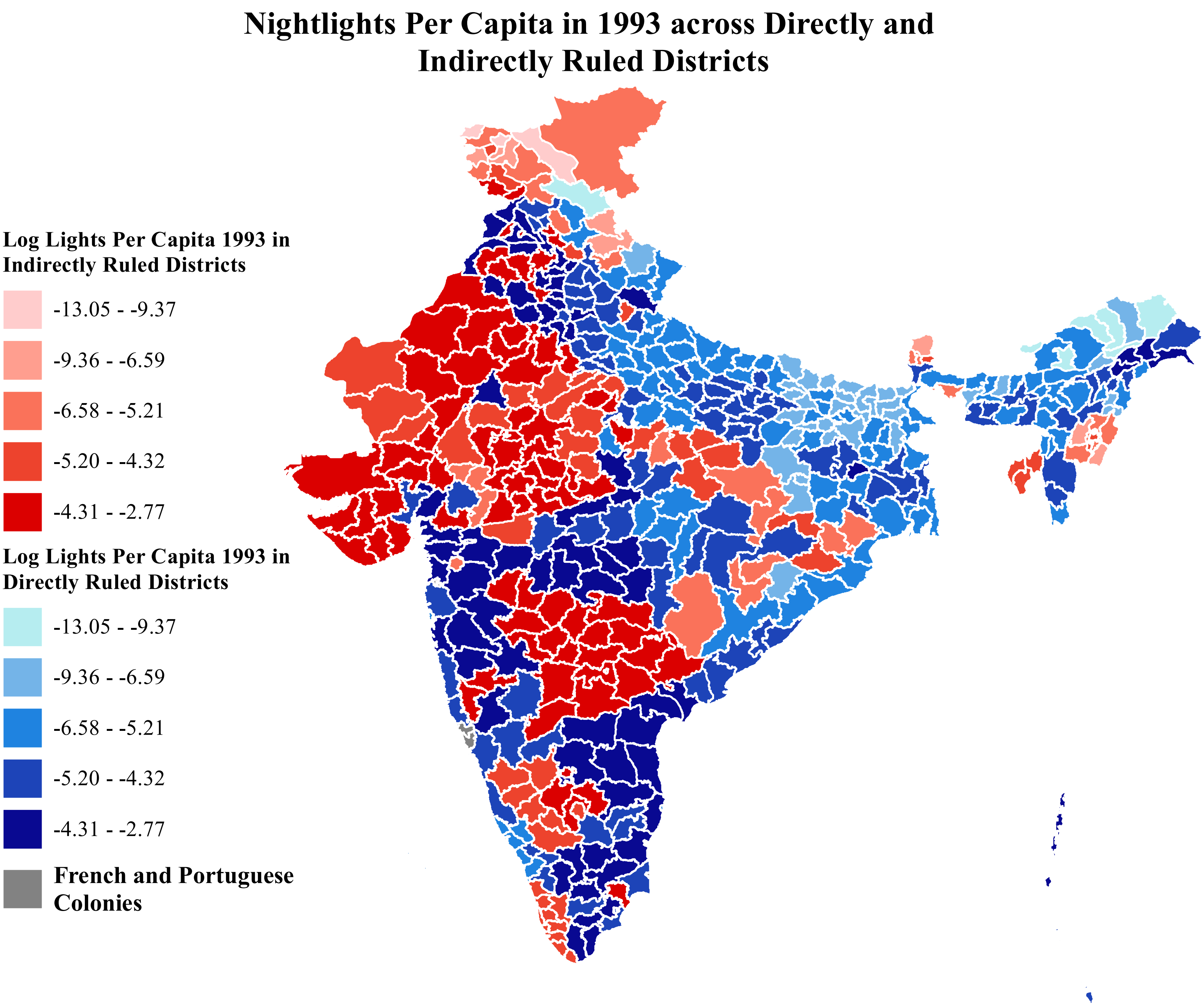

Figure 2: Distribution of Log Lights Per Capita in 1993 across Directly and Indirectly Ruled Districts

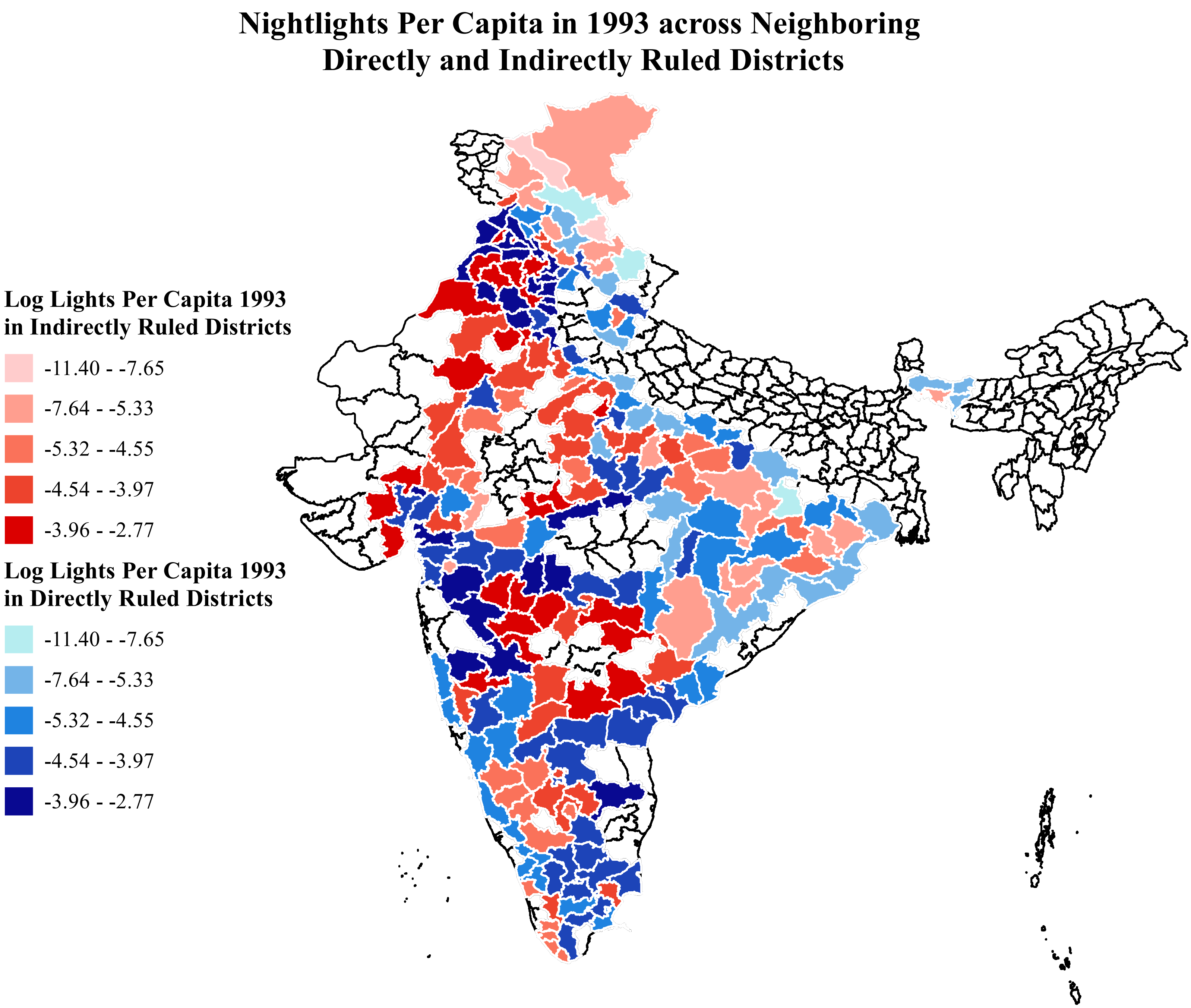

Figure 3: Distribution of Log Lights Per Capita in 1993 across Neighboring Directly and Indirectly Ruled Districts

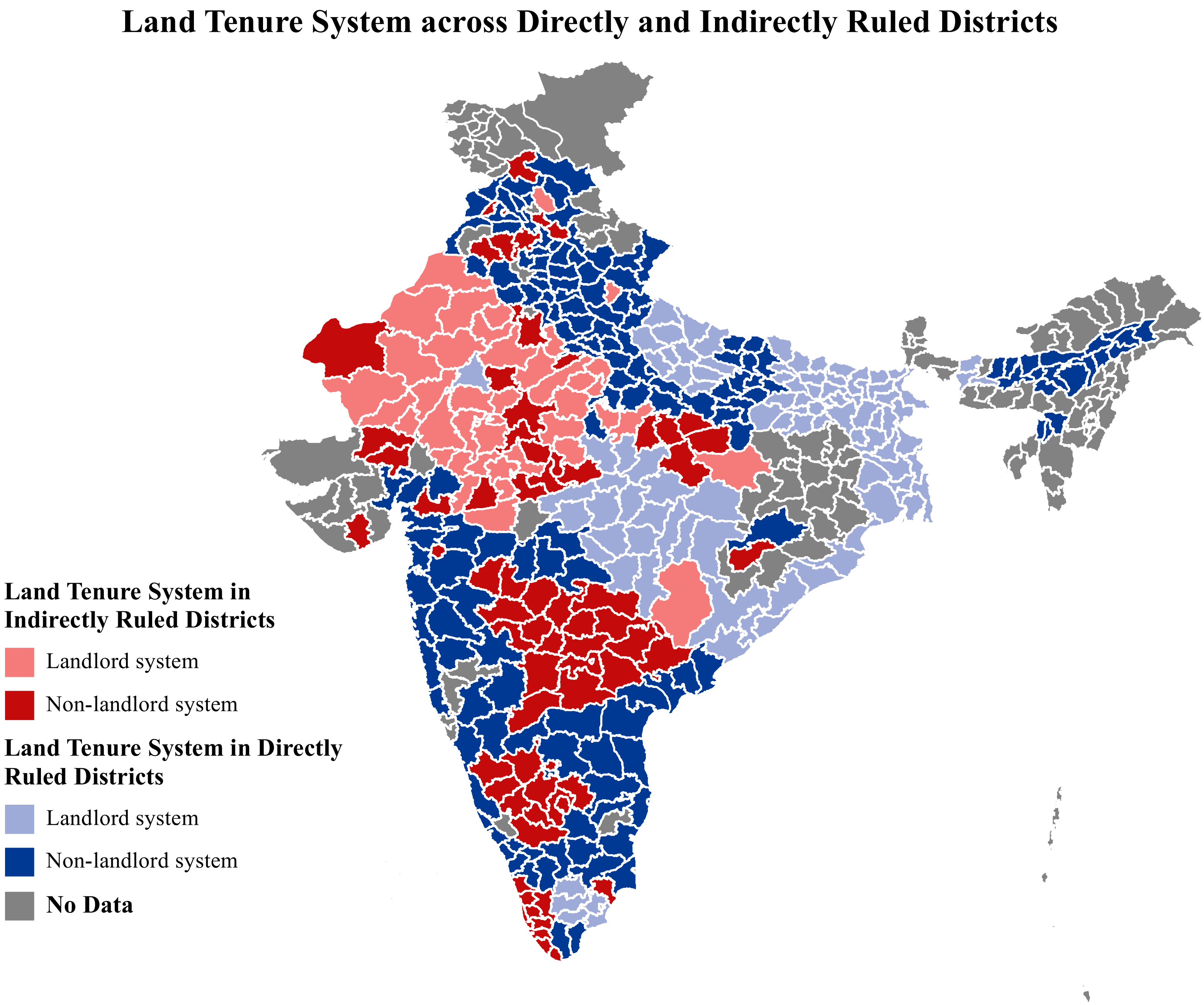

Figure 4: Distribution of Land Tenure Institutions across Directly and Indirectly Ruled Districts

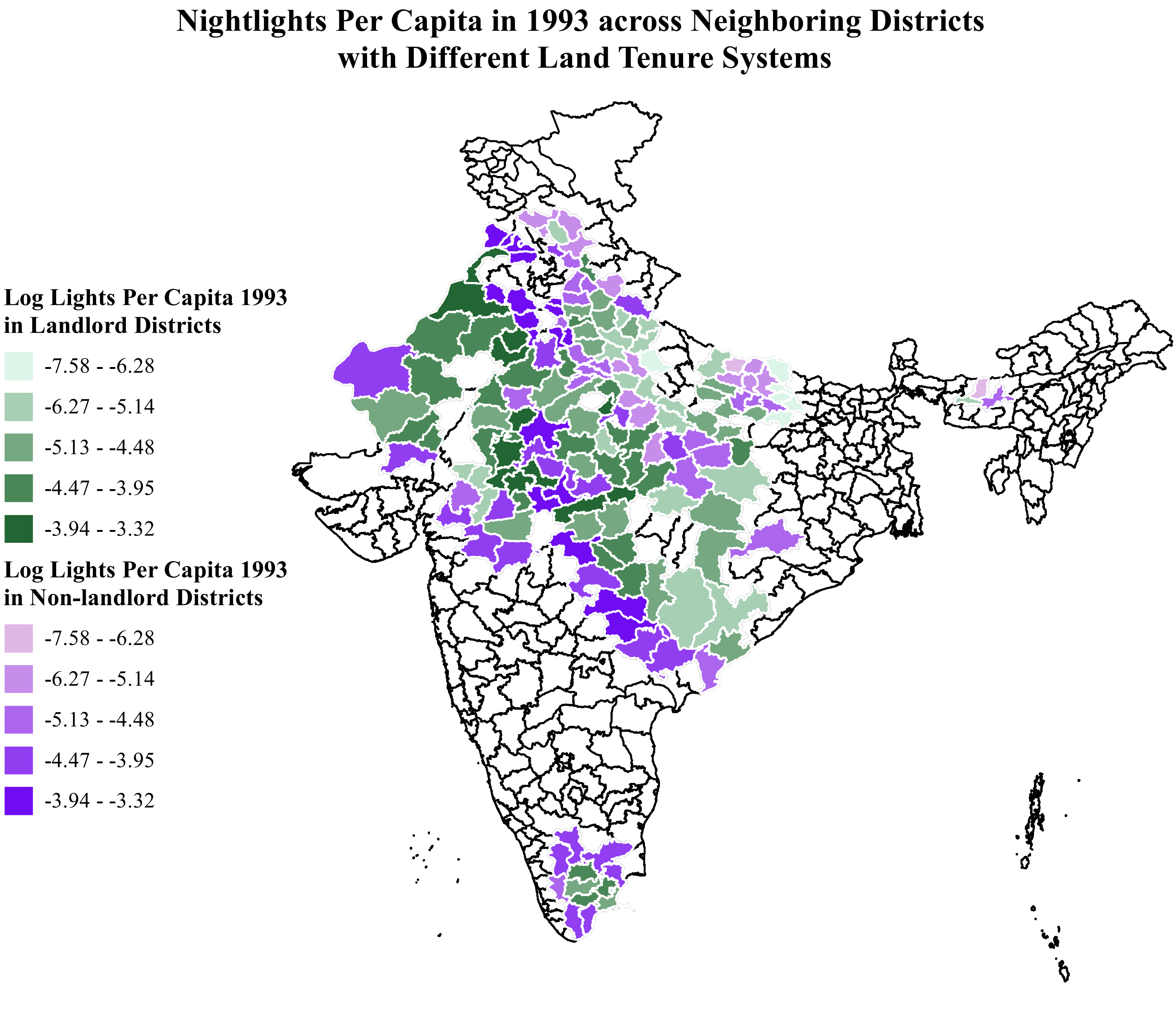

Figure 5: Distribution of Log Lights Per Capita in 1993 across Neighboring Landlord and Non-landlord Districts

Data Appendix

Nightlights data:

Retrieved on June 26, 2020 from https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html. These annual composite images are made up of billions of pixels with each pixel emitting varying intensities of light. These pixels are scaled to a 30 arc-second grid (approx. 1 km^2 at the equator) between 65 degrees south and 75 degrees north latitude. The intensity of light emitted from each pixel is measured using a digital number (DN) that ranges from 0-63. A geographical region with a DN of 0 implies almost no light is being emitted from that particular region. On the other hand, DN of 63 attributes to the highest possible intensity of light that can be emitted from a particular geographical region. To obtain nightlights data at the district level in India, we superimpose the 1993 nightlights composite image with the 1993 district-level shapefile of India obtained from IPUMS[1] International GIS Boundary Files in ArcGIS. The ArcGIS software produces zonal statistics where it adds up the intensity of lights coming from each pixel within each district and provides a *sum of lights* statistic that represents the total luminosity of nightlights emitted in 1993 from each district.

Data on Direct and Indirect British Rule:

While our data on direct and indirect British rule are obtained from Iyer (2010) we made some adjustments in light of some disagreement in the literature about whether a district was under direct rule or indirect rule.[2] We refer to two more recent studies, Verghese (2019) and Castello-Climent et al. (2018), as well as independent sources to validate this. We adjust the classification for 5 districts in Iyer (2010): Solan, Bastar, Balangir, and the Dangs were classified as directly ruled and we reclassified them as indirectly ruled; Chamoli was classified as indirectly ruled and we reclassified it as directly ruled. For 4 of these districts our classification agrees with Castello-Climent et al. (2018) while for the 5th, Solan, we use the same classification as Verghese (2019). In addition to this Iyer (2010) data considers Bongaigaon to be a part of the Goalpara in Assam and Tiruvannamalai Sambuvarayar in Tamil Nadu to be a part of the North Arcot district in Tamil Nadu. However, in 1989, Bongaigaon district (Assam) was formed from parts of the Goalpara and

[1] We provide more information on IPUMS later.
[2] We report all these adjustments to the data in Table A10 in the online appendix.

Kokrajhar districts in Assam. In the same year, North Arcot district was split into Tiruvannamalai Sambuvarayar and North Arcot Ambedkar. We therefore consider these 2 districts (Bongaigon in Assam and Tiruvannamalai Sambuvarayar in Tamil Nadu) as separate observations in our analysis (Kumar and Somanathan, 2009; Law, 2016).

As per the 1991 Census, there were 410 districts across the 17 major states that Iyer (2010) uses in her analysis. However, Iyer (2010) states that there were 415 districts. The reason for this discrepancy is that Iyer (2010) has 17 districts in the state of Punjab (2001 Census) while according to the 1991 Census, Punjab was divided into 12 districts.

For reclassifying districts, we compare historical and modern maps by overlaying historical maps of the political division of India (obtained from Imperial Gazetteer 1909, MapsofIndia (http://www.mapsofindia.com), and David Rumsey Map Collection (India and Farther India – political, 1922)) on modern (1991) district-level map of India in ArcGIS.

Lapse (Instrument for Direct Rule):

This dummy variable is taken from Iyer (2010) with one adjustment. As described above, we changed the classification of Bolangir from directly ruled district to indirectly ruled. The area of the current Bolangir district in Odhisha is comprised of the former Patna state which was a princely state till India's independence. The neighboring Sambalpur state (and the modern Sambalpur district) did come under direct British rule under the *Doctrine of Lapse*. As a result of the re-classification of Bolangir, only 15 districts (as opposed to 16 in Iyer (2010)) came under direct British rule under the *Doctrine of Lapse.*

We further adjust the data for duration of British rule, the native level used for clustering standard errors, and *lapse* (instrument used in 2SLS analysis) data accordingly for these 7 districts as well as 5 additional districts. We report all these adjustments to the data in Table A10 in the online appendix.

IPUMS: IPUMS is an online data inventory containing census data from countries all over the world. The IPUMS project makes data available free of cost for academic research and educational purposes through its website. This project is a collaboration of University of Minnesota, National Statistical Organizations across the world, international data archives and other international organizations.

District-level Population 1991: We obtain district-level population data from the 1991 Census Handbook of India as well as Indian District Database that is overseen by Prof. Vanneman from the Department of Sociology at the University of Maryland (Vanneman and Barnes, 2000). The Indian District Database houses district-level Census as well as agricultural data of India from 1961-1991. Data can be found at http://vanneman.umd.edu/districts/files/index.html.

Natural Earth Data: Natural Earth Data provides free vector and raster map data at 1:10m, 1:50m, and 1:110m scales. This data is organized across three broad categories: cultural, physical, and raster. The coastal boundary shapefile used in our analysis can be found under the physical category at the 1:10m scale.
Data retrieved on Jan 19, 2020 from https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/.

District-level Literacy Rate in 1991: We obtain data on literacy rate in 1991 from the 1991 Census of India. Since no census was performed in Jammu and Kashmir in 1991, we use the average of 1981 and 2001 literacy rates for districts in Jammu and Kashmir instead. The 1991 Census of India Handbook defines *literacy rate* as the percent of literate persons in the age group 7 and above. According to the census, "a person who can read and write with understanding in any language", is considered to be literate.

Historical Location of Catholic Missionaries in 1911: We obtain data on the historical location of Catholic missionaries in 1911 for most districts in our sample from Castello-Climent et al. (2018). 18 modern districts (1991 district boundary) had missing data on historical location of Catholic missionaries in 1911. For these districts with missing data, we overlay the historical map of India published in the first edition of Atlas Hierarchicus by Karl Streit (1913 edition) on the modern district level (1991 level) map of India in ArcGIS to obtain data on historical location of Catholic missionaries in India in 1911.

Infant Mortality Rate (IMR): We refer to Government of India (1997) and manually obtain IMR in 1991 for districts in all states except Jammu & Kashmir (J&K hereon) as no census was held in

J&K in 1991. Alternatively, we refer to India State-Level Disease Burden Initiative Child Mortality Collaborators (2020) and use the IMR in 2000 for the districts in J&K instead.

Road Length Per Capita and Railroad Length Per Capita: We obtain raw data on roads and railroads as of 1992 from DIVA-GIS. We import these vector (line) raw data files for roads and railroads into the ArcGIS software and obtain the total length of roads and railroads (km) passing through each of the 466 districts as of 1992.[3] We then divide these measures with the population of that district to obtain our variables of interest for infrastructure, i.e. road length per capita and railroad length per capita. Similar to nightlights per capita in 1993, we use the natural log of these variables in our analysis.

Institutional Data for excluded districts (45 districts): We manually obtain data for the remaining 45 districts from the respective district government's website as well as the Imperial Gazetteer of India (1909, 1922, 1931). We also verify our coding of these 45 districts by overlaying historical maps of the political division of India on modern (1991) district-level maps of India in ArcGIS. Historical maps were obtained from Imperial Gazetteer 1909, David Rumsey Map Collection (India and Farther India – political, 1922) and MapsofIndia (http://www.mapsofindia.com).

Geographical Controls:
We obtain the area (sq. km) of all the 466 districts in 1991 from the Census Digital Library of India (Office of the Registrar General & Census Commissioner, India) which makes available all Census tables published from 1991-2011. We refer to DIVA-GIS for raw geographic data on elevation and river length and utilize ArcGIS software to obtain the average elevation (m) and total length of all rivers (km) passing through each of the 466 districts. DIVA-GIS provides free spatial data for the whole world. Data include inland water (rivers, canals, and lakes), elevation, administrative boundaries, roads, railroads, and land cover.

We also extract the latitude of the centroid of each district using the ArcGIS software. We follow Chanda and Kabiraj (2020) and obtain the coastal boundary shapefile from Natural Earth

[3] The original source of the roads and railroads data is the Digital Chart of the World (DCW). The DCW is a comprehensive digital map of the world and houses GIS global database on roads, populated places, railroads, utilities, drainage, ocean features, land cover, hypsography, cultural landmarks, etc. This dataset updated by the National Geospatial Agency (NGA) in 1992 and has been freely available for public use since 2006.

Data. We import this into ArcGIS software and compute the distance to the nearest coastline (100 km) from the centroid of each district.

Furthermore, we obtain monthly temperature and precipitation data from the Department of Geography at the University of Delaware which maintains an archive containing gridded monthly time series (1900-2017) of terrestrial air temperature and precipitation data (Willmott and Matsuura, 2001; Shah and Steinberg, 2017). We import this gridded raw data into ArcGIS and obtain average annual temperature (ºC) and average annual rainfall (mm) for each of the 466 districts. We compute the average annual temperature and average annual rainfall for the year 1993 using monthly data from 1900-1993.

Terrain Ruggedness (100 m): We obtain data on terrain ruggedness from Nunn and Puga (2012) and use ArcGIS to extract the index of terrain ruggedness (100 m) for all the 412 districts in our analysis. Data retrieved on Aug 20, 2020 from https://diegopuga.org/data/rugged/#grid.

Agricultural Land Suitability Index: For the agricultural land suitability index, we refer to Ramankutty et al. (2001). Ramankutty et al. (2001) divide the world geographical map into smaller 0.5-degree grid cells and develop an index representing the fraction of land suitable for agriculture in each grid cell using the temperature and soil conditions of each grid cell. We obtain this agricultural suitability index dataset from the Center for Sustainability and the Global Environment, University of Wisconsin – Madison where it is publicly available. We then use the ArcGIS software to extract the average agricultural suitability index for all the 412 districts. Given the nature of the raw dataset, the agricultural suitability index data is only available for 377 districts of the 412 districts in our analysis. Data retrieved on Aug 20, 2020 from https://nelson.wisc.edu/sage/data-and-models/atlas/maps.php.

Historical Controls: We obtain the set of historical controls from Castello-Climent et al. (2018). The set of historical controls include share of urban population in 1931, share of tribal population in 1931, share of brahman (type of caste) population in 1931, presence of railways in 1909, population density in 1931, and number of Europeans in 1931 for the 412 districts in our sample. We match the districts in our dataset to the districts in Castello-Climent et al. (2018) keeping in mind the boundary changes between 1991 and 2001 and are able to obtain these historical controls for 394 districts of the 412 districts in our analysis. The remaining 18 districts are not included in the Castello-Climent et al. (2018) sample and hence we are not able to obtain data on historical controls for them The 18 districts that are not included in the Castello-Climent et al. (2018) sample for which we do not have historical data are: Hyderabad (Andhra Pradesh), Aurangabad (Bihar), Bhojpur (Bihar), Gaya (Bihar), Jehanabad (Bihar), Nawada (Bihar), Patna (Bihar), Rohtas (Bihar), Delhi (Delhi), Lahul & Spiti (Himachal Pradesh), Seoni (Madhya Pradesh), Greater Bombay (Maharastra), Churu (Rajasthan), Chengalpattu M.G.R. (Tamil Nadu), Madras (Tamil Nadu), South Arcot (Tamil Nadu), Calcutta (West Bengal), and West Dinajpur (West Bengal).

<u>Average Consumption Per Capita 1993-1994</u>: Topalova (2010) obtains data on consumer expenditure from the NSS (National Sample Survey) for years 1983, 1987-1988, 1993-1994, and 1999-2000. She calculates the average consumption per capita at the district level in India for each of these years. For our analysis, we obtain the 1993-1994 average consumption per capita data at the district level for 1993-1994 (NSS, 50th round).