

The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java*

Melissa Dell and Benjamin A. Olken

Harvard University and MIT

April 2017

Abstract

Colonial powers typically organized economic activity in the colonies to maximize their economic returns. While the literature has emphasized the negative impacts of colonial extraction on institutional quality, the changes in economic organization implemented to spur production may have countervailing long-run effects. We examine these in the context of the Dutch Cultivation System, the integrated industrial and agricultural system for producing sugar that formed the core of the Dutch colonial enterprise in 19th century Java. We show that areas close to where the Dutch established sugar factories in the mid-19th century are today more industrialized, have better infrastructure, are more educated, and are richer than nearby counterfactual locations that would have been similarly suitable for colonial sugar factories. We also show, using a spatial regression discontinuity design on the catchment areas around each factory, that villages forced to grow sugar cane have more village owned land and also have more schools and substantially higher education levels, both historically and today. The results suggest that the economic structures implemented by colonizers to facilitate production can continue to promote development and economic activity in the long run.

Keywords: long-run development, colonialism, economic organization

*We thank Abdulhamid Haidar, Mateo Montenegro, Roman Zarate and especially Peter Hickman for providing excellent research assistance in Cambridge, Timo Stibbe for assistance in the Netherlands, and Muhammad Abduh for assistance in Indonesia. Contact email: melissadell@fas.harvard.edu, address: Harvard University Department of Economics, Cambridge MA 02138.

1 Introduction

The legacy of extractive colonial institutions for economic development is often thought to be negative. Colonizers whose focus was on extracting resources from the colonies plausibly set up weak institutions with poor property right protections to facilitate this extraction (Acemoglu et al., 2001). These institutions may in turn have persisted, continuing to be a drag on economic performance today.

Counterbalancing this effect, however, is the fact that in many cases colonial powers established complex economic systems to create the economic surplus that they wanted to extract. In the case of agricultural extraction, crops needed to be grown, harvested, and processed before being transported to the home country. This typically involved technology transfer and infrastructure construction, such as the large-scale sugar processing and transport infrastructure that the Japanese created in colonial Taiwan or the tea plantations and processing plants that the British established in colonial India. Extracting a large surplus also often required a substantial reorganization of the indigenous local economy, and these changes likewise may have had long-run local development effects that differ from the overall macro institutional impacts.

In this paper, we examine the long-run local development effects of extractive institutions by studying the Dutch Cultivation System (*Cultuurstelsel*) in Java. The island of Java was the main population center of the vast Dutch colonial empire in the East Indies, and with a modern population of over 160 million, remains the economic and population center of Indonesia today. From the early 1830s through the 1870s, the Dutch colonial state forced peasants along Java's northern coast to cultivate sugar, which was then processed in nearby colonial factories for subsequent export to Europe. The revenues extracted from this system made Java among the world's most lucrative colonies, and at their peak accounted for over one-third of Dutch government revenue (Luiten van Zanden, 2010). Sugar production continued after the abolition of the System but collapsed during the Great Depression, as Indonesia lacked protected markets, and today Indonesia is the world's largest sugar importer.

Prior to the Cultivation System, the Javanese economy was heavily specialized in rice cultivation, largely for subsistence consumption. The System required a substantial reorganization of economic life in order to operate. At its heart were 94 water-powered sugar factories set up by the Dutch, which crushed recently harvested cane and processed it into refined sugar. Over the course of the System, millions of Javanese worked in sugar processing and transport - via both forced and free labor - a major shift since factory production had been nearly non-existent initially (Elson, 1994, p. 215). Since raw cane is heavy and needs to be refined quickly after harvest, the poor infrastructure of the time meant that the cane had to be grown nearby. Accordingly, the Dutch constructed a catchment area with a radius

of approximately four to seven kilometers around each factory, and forced all villages within the catchment area to reorganize their land to grow cane. Local village officials within the catchment areas were empowered by the colonial government to make this happen.

We study the impacts of different aspects of the Cultivation System - the construction of a sugar processing infrastructure as well as forced sugar cultivation - using two complementary identification strategies. We combine these strategies with a wide variety of highly disaggregated modern Indonesian data to isolate the effects of these economic changes.

To estimate the effects of creating a sugar manufacturing infrastructure, we use GIS to construct the locations of alternative, counterfactual sites that would have been equally plausible locations for sugar factories and hence can serve as a comparison group. While there were various similarly suitable locations, catchment areas were often adjacent to one another, and factories could not be too close since each required an adequately sized catchment area. Hence, there were many possible equilibria for site selection (Salop, 1979). We identify counterfactual sites using two criteria. Since state-of-the-art processing technology was water-powered, sugar factories were located along rivers; we therefore only consider sites that can be reached by moving upstream or downstream via river from the actual factory. Moreover, factories needed to have a sufficient amount of sugar-suitable land nearby; we therefore only consider sites for which the amount of suitable land, based on elevation and slope, is similar to the amount of suitable land near actual factories. We then estimate the effects of being close to an actual factory and compute p-values by comparing these effects to the distribution of estimated effects of being near plausible counterfactual sites. To separate out the effects of Dutch extraction from modern sugar industries, we focus on those Dutch colonial factories where there is no modern sugar factory, though most of our results are very similar if we consider the entire sample.

Second, to identify the impact on a village of being subjected to forced sugar cultivation, we use a spatial discontinuity across the borders of the catchment areas. We obtained a 19th century handwritten list of all villages subjected to forced cultivation from Dutch archives in the Hague, and matched these villages with modern georeferenced locations. This allowed us to compute the exact location of the Cultivation System boundaries (see Figure 1). Inside, villages cultivated sugar cane for the government, whereas outside they did not. The boundaries form a multi-dimensional discontinuity in longitude-latitude space and allow us to control for smooth geographic variation, including in proximity to the nearest historical factory. Identification requires that pre-determined characteristics change continuously at the catchment area boundaries, and we provide evidence that this assumption is reasonable.

We find that the Dutch colonial sugar industry substantially transformed economic activity in contemporary Java. SUSENAS household surveys, collected between 2001 and 2011, document that people living within a few kilometers of historical sugar factories are much

less likely to be employed in agriculture and more likely to be employed in manufacturing or retail than people living further away. Effects are similar in Census data from 1980, when Indonesia was more agricultural, suggesting that impacts are not just driven by recent industrialization. A similar analysis using distance to plausible counterfactual locations provides strong evidence that these differences are not driven by geography or random chance.

The composition of industries in these areas suggests a channel of persistence. Using national input-output tables, we classify manufacturing employment in the 2006 Economic Census into employment that is upstream and downstream from sugar processing. Upstream manufacturing industries - those whose outputs sugar processing uses directly or indirectly as inputs - include farm and capital machinery. Downstream industries - those that use processed sugar as an input - include most other food processing industries. We find more employment both upstream and downstream in areas near historical factories. When we restrict to those historical locations with no modern sugar factory, the upstream effect disappears, but we still find substantially more employment near historical sugar factories in downstream industries. This suggests a path for persistence. While the Dutch government claimed the high quality processed sugar, the factories were permitted to sell the low quality sugar, which was costly to transport, on local markets. Industries that used sugar as an input sprung up surrounding colonial sugar factories, creating food processing centers. Even when the original source of sugar disappeared, those industrial centers persisted, sourcing sugar from elsewhere. Food processing, as in many developing countries, is central to manufacturing in Indonesia today.

Beyond constructing sugar factories, the Dutch also built road and rail infrastructure to transport processed sugar to the ports, and this infrastructure has persisted through the present. Villages located within a few kilometers of a historical sugar factory are more likely to have a paved road in 1980 and have a much higher road density today. Public goods more generally are also higher. Villages near a historical factory are much more likely to have electricity in 1980; and are more likely to have had a high school in 1980, when high schools were very rare. Using the complete 100% sample microdata from the 2000 Census, we find that people in villages within a few kilometers of colonial factories are more likely to be educated. This is true both for cohorts born in the 1920s, who completed their education during the Dutch era, and for more recent cohorts who completed their education under an independent Indonesia. Moreover, villages near a colonial sugar factory are closer to modern subdistrict (*kecamatan*) capitals.

In summary, there is clear evidence that they are more developed in a variety of ways. Consistent with this, we also find they are richer. Modern household survey data show that households living within a few kilometers of a historical factory location have per-capita consumption that is about 15 percent higher than those living more than 10 kilometers away

from a historical factory.

The discussion thus far has focused on one aspect of the Cultivation System, namely the establishment of the factory itself. The second part of our analysis turns to the villages that were forced to grow sugar cane. We estimate the impacts on these villages using a spatial discontinuity design, comparing them to nearby villages just outside the Cultivation System catchment areas.

The historical literature suggests that the Dutch relied substantially on village heads to implement cane production and gave them considerably greater command over land and labor than they had exercised previously. In particular, village heads were tasked with choosing the location of the cane fields - which had to be rotated each year - and distributing the remaining land amongst private village landholders and public village land (Elson, 1994).

We therefore begin by examining how this process changed the modern distribution of land, focusing on two types of land. First, Javanese villages typically have some amount of land set aside permanently for use by the village for public purposes. Temporary use rights over some of this land (known as *tanah bengkok*) is given to village heads and other village officials as compensation for their service; revenues from the use of other land enters the village treasury for other purposes. We find that there is about 10 to 15 percent more of this type of public-use land in Cultivation System villages. This has remained fairly constant over time, with virtually identical effects in both the 1980 and 2003 datasets. Second, we examine the impact on the distribution of private agricultural land. Effects on private land inequality are if anything positive but fall short of statistical significance.

We then turn to economic structure. We find that households in Cultivation System villages appear more educated. Echoing the previous results, educational effects for households in these villages go all the way back to the 1920s cohort, which was educated during the Dutch period. During the early 20th century, villages that wanted a school needed to fund the school building themselves. The revenue from the extra village land in these villages may have been used to fund school construction, and indeed we find evidence that these places had more school infrastructure prior to the nationally-funded school construction program begun in the 1970s and studied by Duflo (2001). The village heads, who receive more land as compensation, are also more highly educated.

These villages also continue to have a different economic structure, with a greater percentage of households working in manufacturing and retail and fewer in agriculture, both in 1980 and today. They also have more manufacturing firms on average than villages located just outside the discontinuity, and are more densely populated. People living in these villages are no poorer than those living elsewhere. Since we control flexibly for distance to the historical factory and geographic location, these effects isolate the impacts of being subjected to forced cultivation, and are thus above and beyond the effects of being closer to a factory,

identified above.

In sum, we find that the economic changes induced by the Cultivation System led to a more modern economy, with more manufacturing and retail and greater education levels – both for the areas in the immediate vicinity of the historical sugar factories and further out in the villages where the economy was reorganized to grow sugar cane.

These results inform three different debates about the historical origins of modern economic structure. First, they speak to a classic literature on the role of input-output linkages in economic development. During the 1950s and 60s, development economists hypothesized that linkages between sectors provided a key mechanism for the propagation of structural change and debated the merits of targeting industrial policy based on the density of input-output linkages. (Rasmussen, 1956; Myrdal, 1957; Hirschman, 1960). More recent empirical evidence documents the importance of input-output linkages in transmitting macroeconomic shocks (Acemoglu et al., 2016; Baqaee, 2015; Acemoglu et al., 2012; Carvalho, 2009) and in development (Liu, 2017). However, there is little empirical evidence on the role that they play in the long run. The Cultivation System provided local markets with low grade sugar that was costly to transport, generating benefits from agglomeration with downstream producers. Relative to other agricultural products, sugar has very dense linkages, since it is added to most processed foods. We find substantial effects on downstream firms that persisted even after the original sugar factories disappeared, perhaps because of input-output linkages amongst remaining firms and the endogenous concentration of final goods consumers near initial production sites.

Second, to the extent that not all returns from village lands are extracted for private consumption, the persistently larger public village landholdings in Cultivation System villages provided the local government with additional ability to raise revenue, a classic measure of state capacity, which we might expect to lead to increased public goods provision. A classic literature (Hayami, 1980; Scott, 1977), as well as more recent empirical evidence (Dell et al., 2017), argue for the importance of village level organization and public goods provision for long-run economic development. Our results suggest that village government is indeed important, even as the central state has become increasingly involved in public goods provision in recent years.

Third, the study contributes to an extensive literature on the long-run impacts of extractive colonial institutions and relates to a large body of work that highlights the relevance of historical institutions and persistent cultural traits for long-run development.¹ In a series of

¹See Becker et al. (2016); Guiso et al. (2016); Lowes et al. (2015); Acemoglu et al. (2015); Bukowski (2016); Oto-Peralías and Romero-Ávila (2014); Grosjean (2014); Michalopoulos and Papaioannou (2014); Acemoglu and Robinson (2013); Spolaore and Wacziarg (2013); Michalopoulos and Papaioannou (2013); Alesina et al. (2013); Voigtländer and Voth (2012); Acemoglu et al. (2011); Nunn and Wantchekon (2011); Luttmmer and Singhal (2011); Grosjean (2011); Tabellini (2010); Fernández and Fogli (2009); Tabellini (2008);

innovative cross-country studies, Acemoglu et al. (2002, 2001) hypothesize that extractive colonial institutions are at the heart of much of the developing world’s poor long-run growth performance. A subsequent literature aimed to refine identification and isolate channels of persistence by focusing on within country variation, with several studies examining forced labor of various types. In particular, Dell (2010) uses a spatial discontinuity to document that forced labor in the Potosi silver mines during Spanish colonization lowered economic prosperity in the long-run and hypothesizes land tenure and public goods provision as channels of persistence, a finding echoed in an examination of forced rubber cultivation in the Congo by Lowes and Montero (2016).

While the economics literature has focused on the detrimental long-run effects of extractive colonial institutions, qualitative historical accounts of major instances of colonial extraction highlight that reality was often complex. Colonialism plausibly had macro institutional effects, but also had major effects on the local organization of economic production. Colonial states frequently extracted revenues via forced labor (Van Waijenburg, 2015), but *how* these institutions impacted the local economy varied considerably. In some cases, the forced labor took place within an individual’s community, whereas in other instances, individuals were removed from their communities and forced to labor in a single location hundreds of miles away. Moreover in some instances, the nature of the technology required industrial processing to occur where the output was produced, whereas in others, the raw materials were exported and processed elsewhere. Different types of extraction involved varying levels of technology transfer, with varying degrees of linkages to other economic sectors. This study documents how in-situ changes in production, combined with nearby industrial processing, had a positive effect on long-run development in Java. The next section provides evidence that the Dutch in Java were unlikely to be unique in implementing a colonial economic system that persistently spurred economic output and encouraged a more modern economy.

The remainder of the study is organized as follows. Section 2 discusses the historical context, and Section 3 describes the data. Section 4 examines the long-run impacts of the sugar processing infrastructure established by the Cultivation System, whereas Section 5 estimates the effects of being subjected to forced cultivation. Section 6 concludes.

2 Historical Background

2.1 Historical Overview

The Cultivation System (*Cultuurstelsel*), in force from the early 1830s through the 1870s, compelled Javanese villagers to produce export crops for the Dutch colonial government. At

Nunn (2008); Guiso et al. (2008); Gennaioli and Rainer (2007); Giuliano (2007); Banerjee and Iyer (2005).

its peak, the Cultivation System provided over one third of Dutch government revenues and four percent of Dutch GDP (Luiten van Zanden, 2010; Ricklefs, 2008, p. 159), making Java one of the world’s most financially lucrative colonies. While a variety of crops were grown, from the 1850s onward sugar and coffee accounted for more than 96% of profits (Elson, 1994, p. 135).² We focus on sugar cultivation, which took place on the populated plains of Java’s Northeast Coast. Coffee, while profitable, was primarily grown on remote forested mountain slopes that were uninhabited initially (Elson, 1994, p. 65). Extant data on coffee cultivation are at a higher level of aggregation that do not permit the detailed analyses that we perform for sugar.

The sugar system was a major enterprise, as Java was the world’s second largest sugar producer at the time, surpassed narrowly only by Cuba. The system started in the early 1830s with over 70,000 sugar growers cultivating cane for 59 factories, and over time grew to encompass 94 factories and a large share of the Javanese population (Elson, 1994, p. 55).³ A Dutch report cited by Fasseur (1992) notes that as much as 25 percent of the native Javanese population was involved in the Cultivation System (excluding coffee), of which sugar was the largest component. Dutch reports from the 1860s show that over 2.5 million forced workers labored in the sugar factories or related services (i.e. transporting cane and firewood to the factories), joined by free laborers whose numbers expanded significantly across time (Elson, 1994, p. 215). Sugar factories were run by private entrepreneurs with close links to the Crown, who were required to sell their high quality sugar to the government. Villages surrounding each factory had to allocate up to one-fifth of their land to grow cane, though in practice this ratio was often significantly exceeded (Elson, 1994, p. 229; Van Niel, 1992, p. 137). Europeans were prohibited from renting or purchasing land and from establishing their own private factories, and hence sugar production occurred almost exclusively through the Cultivation System.⁴

The Cultivation System was primarily administered by Javanese officials, who gained considerably greater command over land and labor than they had exercised previously (Van Niel, 2005, Elson, 1994, p. 183). Village heads were in charge of allocating sugar cultivation land and with assigning labor to cultivation services. The village received incentive payments (*kultuurporcenten*) for cultivating cane, and the village head was also tasked with distributing these and with collecting the land tax owed to the Dutch government.

The Dutch phased out the Cultivation System in the 1880s, and sugar expanded rapidly

²Indigo was grown initially but the Dutch started phasing it out in 1834 due to low profitability, and indigo cultivation declined sharply throughout the 1840s (Elson, 1994, p. 83, 110, 131; Van Niel, 1992, p. 112). Cinnamon, tea, and tobacco were grown by a small number of peasants in localized areas.

³The factories were primarily established in the 1830s and after that changed little. (Elson, 1984, p. 71).

⁴The exceptions were the indirectly ruled Principalities, which had private estate agriculture and were exempt from the Cultivation System.

under free enterprise (Elson, 1984, p. 131). However, the industry collapsed during the Great Depression, and subsequently Java's primary competitors enjoyed large protected markets, whereas Java did not.⁵ Moreover, much of Europe and North America subsidized and protected new domestic beet sugar producers. Today Indonesia is a major sugar importer, with sugar contributing only 0.05% of their agricultural exports.

Government sugar cultivation exerted a variety of impacts, on subjected villages as well as on villages in proximity to sugar factories. An older literature argues that the System led to an equalizing of the land distribution, as villagers aimed to share the System's burdens (Geertz, 1963). This equalization is hypothesized to have stifled modernization by preventing the emergence of a class of prosperous entrepreneurial farmers. A subsequent literature likewise hypothesizes that the Cultivation System had equalizing effects, by increasing the amount of communal land held by the village, by leading communal land to be divided more equally, or by increasing the share of landowners.⁶

A revisionist literature, initiated by Elson (1994, 1984), argues instead that the Cultivation System empowered Javanese village elites, increasing hierarchy within the village. There were only a handful of Dutch officials in Java, concentrated in major cities, so the Dutch relied on Javanese authorities to ensure that resources were devoted to government cultivation. In order to guarantee their cooperation, the Dutch made village authorities accountable to the colonial state and ensured that they benefited from furthering Dutch objectives.⁷ Village heads were responsible for assigning land for government cultivation - which had to be rotated on a regular basis - and they often used this as a pretext to redistribute land towards themselves and their cronies (Van Niel, 1992, p. 139). The Dutch could ensure their cooperation with the sugar system by permitting such behavior. By 1857, 20% of all the irrigated land in Probolinggo Residency - a center of sugar cultivation - was held by officials, in Kedu village chiefs had taken over more than half of the land, and similar patterns obtained in Pekalongan-Tegal and other sugar-intensive areas (Elson, 1984, p. 94). Ricklefs (2008, p. 159) argues that while communal village landholdings increased in some areas as a response to the System, instead of having a leveling effect this likewise empowered the village elite, who were in charge of allocating village land.

Despite these distributional effects, historians such as Elson (1994) have argued that the Cultivation System positively impacted economic outcomes for the typical villager. Subjected communities received incentive payments, based on the amount of sugar produced,

⁵For example, the Philippines and Hawaii had protected access to the U.S., and Taiwan had protected access to Japan.

⁶See Elson (1994, p. 162) for a review.

⁷Traditionally, the head's term was limited and required consent from landholders. Dutch officials increasingly intervened to install village authorities friendly to their interests and resisted the periodic rotation of village heads (Ricklefs, 2008, p. 158; Elson, 1994, p. 172-174). Following an 1854 reform, succession of the village head was made hereditary.

which often but not always exceeded the land tax owed to the Dutch (Elson, 1994, p. 311, Van Niel, 1992). In places without forced cultivation, money for taxes had to be obtained by selling rice to Chinese traders, who colluded with officials to monopolize the rice trade. Money from cultivation payments allowed citizens to avoid selling large quantities of rice at low prices just before the land tax came due. Historians have also argued that crop payments, by injecting currency into the economy, promoted modern exchange (Elson, 1994, p. 261, Ricklefs, 2008, p. 158). Residencies (provinces) more involved in the Cultivation System experienced a greater growth in markets while the System was in force.

Historians have also pointed to modernizing economic effects that depend on proximity to the sugar processing centers. Extensive road and rail infrastructure were constructed to connect sugar producing regions to ports, plausibly promoting market integration (Elson, 1994, p. 251-252, Ricklefs, 2008, p. 158). Moreover, the Cultivation System increased economic specialization by creating a variety of economic opportunities outside forced cultivation (Elson, 1994, p. 207-208). Factory and transport workers were often paid a wage, and there were also employment opportunities making baskets and matting to transport the cane, making clay pots for use in the factories, and transporting firewood to the factories. Moreover, while the high grade refined sugar was exported, the low grade sugar could not be exported due to a high water content. The factories were permitted to sell this sugar locally, potentially spurring other food processing establishments to locate nearby.

The Dutch were not the only colonial power that attempted to institute major economic changes in order to extract more surplus. The Japanese in Taiwan are a classic example of a developmental colonial state that made massive investments in increasing output in order to maximize extraction, including the construction of a large-scale, modern sugar-processing infrastructure. The establishment of tea plantations and a network of tea processing plants in India - including the formation of the Assam Tea Company under the auspices of the colonial British government - likewise bears similarities to the Cultivation System, as does the establishment of a colonial Indian jute production and processing infrastructure (Tomlinson, 2014). Van Waijenburg (2015) documents the widespread nature of labor conscription by colonizers in Africa, highlighting many different types of economic arrangements including some that involved labor in agricultural export and processing enterprises.

2.2 Assignment to Sugar Cultivation

Understanding why sugar factories were located in particular places and why certain villages were subjected is central to identifying the long-run impacts of the Cultivation System. During the 18th century, sugar cultivation in Java was limited to a small number of estates in the hinterland surrounding Batavia (now Jakarta), the capital of the Netherlands Indies

and location of most European settlement. These estates crushed cane using wooden rollers pulled by water buffalo and the output was consumed in local markets. During the 19th century, Batavian cane production nearly disappeared, as the technology was primitive and the land marginally suitable (Van Niel, 2005, p. 133, 139).

Dutch official correspondence - analyzed by van Schaik (1986, p. 183) - highlights several factors used to locate sugar factories. Sugar grows well in plains and valleys and does not grow well on mountain slopes, and factories had to be near locations suitable for cultivation since raw cane was costly to transport. However, sugar was not grown in all agronomically suitable locations, as illustrated by the large-scale expansion of cultivation following the end of the government system. Proximity to water was crucial because sugar processing technology in 1830 used iron crushing cylinders powered by water wheels located along rivers. It was not until the late 19th century that steam power became dominant. Proximity to timber was also important but is unobserved, since rapid deforestation occurred throughout the 19th century and forest maps prior to the Cultivation System have not survived. Finally, there needed to be nearby population to cultivate the cane and the factories could not be located in areas where transport costs were prohibitive; hence, the remote valleys of southern Java were not suitable, despite their agronomic conditions. The government factories were created in the 1830s, and were relatively fixed for the remainder of the period.

Each factory was allocated a certain area from which to draw the land and labor it required. Villages were typically within four to seven kilometers of their factory. The Dutch referred to this as the ‘circle system’. Catchment areas tended to be adjacent to each other and avoided major pre-existing cities, which primarily consisted of residency (province) capitals.⁸ Once one factory was sited, another one could not be located too close even if there were many suitable places along a river, since each factory required a sufficiently sized catchment area. We will exploit this fact to create suitable counterfactual factory locations that can be used as a comparison group.

3 Data

3.1 The Cultivation System

Data on the Cultivation System are drawn from manuscript archival records held by the Hague (Commissie Umbgrove, 1858). Constitutional reforms in the Netherlands in 1848 placed colonial affairs under partial parliamentary control, and the Minister of Colonies

⁸Colonial Java was divided into 24 residencies, which were the center of colonial administration and the major cities of the time. Modern Indonesian provinces are larger, with only 6 provinces on Java, but the colonial residency capitals by and large remain the largest cities today.

subsequently ordered an inquiry into government sugar cultivation. The commission was appointed in 1853 and spent four years collecting detailed data. Handwritten documents list which villages contributed to each sugar factory and how much land and labor each village provided. They also contain extensive qualitative information about the social situation in sugar producing areas and sometimes show sketches of the catchment areas.

Historical villages are matched with coordinates from the U.S. National Geospatial Intelligence Agency’s extensive Geonames database, which provides a detailed list of populated places in Indonesia, including many sub-village units. Matches are made using the historical village’s name and the location of its historical district.⁹ The manuscripts also list the distance between the village’s fields and the factory, which allows us to distinguish between multiple matches and provides an additional check. We match 6,383 historical villages with Geonames coordinates, which are located in 2,519 modern village polygons. Villages are small - there are over 30,000 in Java today - and hence the monographs allow us to know the precise location of government sugar cultivation. The sugar factories are matched using the same procedure, and the combination of the historical sub-village, plus the fact that the factory had to be adjacent to a river, means that the factories can all be precisely located.

Not all historical sugar villages can be matched, as some disappeared or changed their names, leaving gaps in the matched catchment areas. The actual catchment areas were contiguous, and hence we infer a contiguous set of subjected villages by drawing straight lines from each sugar factory to the coordinates of contributing villages, assigning any unmatched village along the lines as treated. This method is illustrated in Figure 2, which shows the factories, matched village points, straight lines, and catchment boundaries for an example catchment area. Modern village polygons are shown in the background. 82% of modern villages in the constructed catchment areas are matched to at least one village in the historical dataset. We designate all villages inside these catchment areas as treated. The appendix shows that results are highly robust to instead designating matched villages as treated and instrumenting these with being inside the catchment areas.¹⁰

3.2 Outcome Data

Outcome data are drawn primarily from population, industrial, agricultural, and village censuses and household surveys. These data were collected between 1980 and the present by the Indonesian government’s Central Bureau of Statistics (BPS) and are described in more detail in the data appendix. All of these datasets contain village identifiers that allow us to

⁹Data are available for download at <http://geonames.nga.mil/gns/html/index.html>.

¹⁰The first stage is very strong, with an F-stat of around 1,450 when standard errors are clustered at the sub-district level, making these estimates similar by construction.

match the microdata to modern village boundaries.¹¹

The variation that we exploit to identify the impacts of being a subjected village and being near a historical factory is extremely fine. Anything above the village level aggregates away the relevant variation, and villages are very small, with over 30,000 of them in Java. While a number of older data sources are available, they are too aggregated to be useful for the analysis. We can, however, trace some effects further back in time, using cohort analysis from more recent datasets.

4 Impacts of Sugar Factories

4.1 Empirical Strategy

A central part of the Cultivation System was the establishment of a network of local sugar processing plants, as raw cane could not be transported long distances. These plants may in turn have influenced economic activity in the long run, through agglomeration or other channels, even though most of them have long since disappeared.

Estimating the effects of proximity to a historical factory requires identifying a plausible counterfactual, since pre-characteristics could vary with distance to areas suitable for factories. We exploit the fact that there are a variety of points upstream and downstream from the factories that are similar in terms of water flow and the sugar suitability of surrounding areas, but most did not have a factory. Sugar catchment areas were typically adjacent and each needed to produce enough cane for the factory to operate, so factories had to be spaced several kilometers apart. This suggests that there were many possible equilibria for where factories could have been located (Salop, 1979).

To assess the long-run impacts of proximity to historical sugar factories, we conduct the following counterfactual exercise. For each historical factory, we take 1000 independent random draws that shift the factory anywhere from 5 to 20 kilometers to a suitable counterfactual location along the river. To be suitable, the point must have at least as much sugar suitable land within a 5 kilometer radius as the 10th percentile of the actual factory distribution. Sugar grows in the plains and not on steep hillsides or in mountainous areas. We infer suitability by observing where sugar was grown historically, first computing the 90th percentile of slope and elevation in sugar villages and then assigning anywhere with slope or elevation less than these cutoffs as suitable. We choose the 90th percentile because some sugar villages contain both plains and part of a mountain, leading the right tail of the

¹¹Note that by “village”, we refer to the lowest administrative level in Indonesia. In rural areas these are known as *desa*; in more urban areas, these are known as *kelurahan*. We do not distinguish between *desa* and *kelurahan* in the analysis, and they are treated identically in modern Indonesian data.

elevation and slope distributions to be highly skewed.¹²

This approach is illustrated in Figure 3. Panel a) shows an actual factory, surrounded by a 5 kilometer radius, with sugar suitability and rivers in the background. Panel b) shows a suitable counterfactual factory location, which has a sufficient amount of nearby sugar suitable land, as compared to the distribution of actual factories. Panel c) shows a larger set of counterfactual factories.

We then estimate the following specification, both for the actual factory and for the 1000 sets of counterfactual factory locations:

$$out_v = \alpha + \sum_{i=1}^{20} \gamma_i dfact_v^i + \beta \mathbf{X}_v + \sum_{j=1}^n fact_v^j + \epsilon_v \quad (1)$$

where out_v is an outcome of interest in village v , and the $dfact_v^i$ are indicators equal to one if village v is 0-1 km from the nearest (placebo) factory, 1-2 km,..., 18-19 km. The omitted bin is 19-20 km, which is the maximum distance included in the sample. $\mathbf{X}_v$ are geographic controls for elevation, slope, distance to the coast, distance to the closest natural harbor, distance to the nearest river, log flow accumulation, and a spline in distance to the nearest 1830 residency capital, with kink points every three kilometers. The $fact_v^j$ are nearest factory fixed effects, which ensure that villages are compared to other villages near the same (placebo) factory.

We use the position of the $dfact_v^i$ coefficients for the actual factory in the absolute value distribution of the 1000 counterfactual $dfact_v^i$ coefficients to compute a p-value. Small p-values imply that patterns near the actual factories would have been unlikely to arise in the factories' absence. We plot the coefficients for proximity to the actual factory, denoting significance using the counterfactual distribution. We also document the general shape of the relationship by plotting a linear spline with kink points every 3km that is fitted using the raw outcome data and controls for geographic characteristics and factory fixed effects. Our main set of results restricts the sample to historical factories that are not in close proximity to a modern factory, to ensure that effects are not primarily driven by the persistence of sugar processing itself, and we discuss explicitly the few cases where this restriction matters (primarily: manufacturing of products used as inputs to sugar processing).

The study also conducts an additional counterfactual exercise, shifting all the historical factories up or down the river by the same distance. Then it estimates equation (1) for each of these shifts. We would expect the γ_i to be largest for shifts around 0 km - which do not substantially change the location of the factories - and to dissipate the further the placebo factories are moved away from the actual factories.

¹²An alternative would be to use FAO suitability data, but unfortunately these are too aggregated to be useful when exploiting village level variation.

In our baseline analysis, we also limit the sample to sugar suitable villages, as the analysis of the impacts of forced cultivation is limited to these villages and we want to use a consistent sample throughout the study. The appendix shows that estimates are very similar when the analysis is not limited to sugar suitable villages. The sample also excludes cities that were pre-period residency (province) capitals, which were Java’s major historical cities and were exempt from the Cultivation System since cities are not suitable for agriculture, but again estimates are similar when these are included.

Figure 4 documents that a variety of geographic pre-characteristics vary similarly with distance to the actual and counterfactual factories, suggesting that the counterfactual locations were indeed similar to the actual ones prior to the construction of the factories. The characteristics considered are elevation, slope, distance to the coast, flow accumulation, distance to the nearest natural harbor, and distance to the nearest river. Flow accumulation is a measure constructed by the USGS Hydrosheds project that calculates how many cells are uphill from the cell under question. The higher the number, the more water we would expect to flow through the cell. While a few coefficients are significant, most likely by chance, there is no consistent pattern. Moreover, the magnitudes are very small – for example, moving from 1km away from a factory location to 10km away from a factory is associated with a difference in elevation of about 2 meters – suggesting that these locations are indeed quite similar.

Places near historical factories are closer to 1830 residency capitals (panel g). There are a small number of these - only 24 in all of Java and Madura - and the appendix shows that dropping factories within 10 kilometers of a residency capital flattens out this relationship, whereas our other results are robust to this variation. When specific instead of random shifts are used to compute p-values, moreover, these impacts are not significant in comparison to the placebo distributions, as also documented in the appendix. Finally, we control for a flexible spline in distance to the nearest 1830 residency capital throughout the analysis.

4.1.1 Overall economic structure

To the extent that agglomeration economies are important and highly persistent, government sugar factories could plausibly influence industrialization in the long-run, even in places where sugar has not been processed for many decades. We examine this using a variety of data on individuals’ sector of employment. We focus here on places where there is not a modern sugar factory nearby, in order to isolate effects that do not go directly through sugar, and subsequently we will explicitly examine modern sugar production.

We begin by illustrating the methodology in more detail, using as an example dependent variable an indicator for whether the individual works in the agricultural sector, taken from

the SUSENAS 2001-2011 household surveys.¹³ The sample includes prime age males aged 18 to 55 to avoid confounding labor market participation (estimates are similar using all individuals). As described in Section 4.1, for each factory we take 1,000 independent random draws that shift the factory to a suitable counterfactual location along the river. We estimate equation (1) using distance to the actual factories, as well as the distance to each of these 1,000 sets of counterfactual factory locations.

The sub-plots in figure 5, panel a) show the counterfactual distributions of absolute coefficients for each of the bins in distance to the nearest factory: 0-1 km, 1-2 km, etc. The coefficients that measure the impacts of proximity to *actual* factories are denoted by a red line. Each sub-plot also reports a p-value, given by the fraction of the absolute value counterfactual coefficients to the right of the absolute value of the actual coefficient. For bins near a factory, the actual coefficients fall far in the tails of the counterfactual distributions, indicating that patterns of agricultural employment near government factories would have been very unlikely to arise in the factories' absence. For bins further away, the actual estimates fall in the center of the counterfactual distributions, documenting that agricultural employment is not different from what we would have expected in the factories' absence. Effects dissipate within five kilometers.

Panel b) plots the $dfact^j$ coefficients for proximity to the actual factories. Crosses indicate coefficients that are above the 95th percentile of the counterfactual distributions shown in panel a), solid dots denote coefficients that are above the 90th percentile, and hollow dots indicate coefficients that are below the 90th percentile. These figures also plot a linear spline with kink points every 3 kilometers. The spline is fitted on the raw outcome data, controlling for geographic characteristics, nearest factory fixed effects, and survey year fixed effects. Individuals within a few kilometers of a historical factory are around 20 to 25 percentage points less likely to work in agriculture than those 10 to 20 kilometers away. The appendix shows a version of these plots where the mean of the placebo distribution has been subtracted out from each coefficient, and impacts remain similar.

We also conduct a placebo exercise that shifts all the historical factories up or down the river by the same distance. Equation (1) is estimated for each of these shifts. Each sub-plot in panel c) shows the γ_i coefficients for a bin in distance to the nearest factory: 0-1 km, 1-2 km, etc. The x-axis plots the magnitude of the shift upstream (left side of the plot) or downstream (right side of the plot), in kilometers. The y-axis plots the $dfact^j$, with the length of each bar indicating the magnitude of the counterfactual estimate. For villages near the alternative factory locations, the γ_i are largest for shifts around 0 km,

¹³When we use the SUSENAS datasets, for which we pool a number of different years of the data together in order to have sufficient number of observations in each village, we augment equation (1) to include a survey-year dummy.

which do not substantially change the location of the factories. Effects dissipate the further the counterfactual locations are from the actual ones. For bins located further from the factories, there is little effect regardless of the magnitude of the shift, since further places are not influenced by the actual factories. Panel d) is analogous to panel b), plotting the coefficients for proximity to the actual factories and computing p-values using counterfactual distributions constructed with the specific distance shifts.

Figure 6 examines the industrial structure near government sugar factories in more detail, examining both different sectors and different time periods. Due to space constraints, for the remainder of the paper, only the coefficient plots are shown in the main text, as well as the statistical significance for each γ_i coefficient estimated using the counterfactual distributions from the independent random shifts. The complete set of counterfactual distributions for the random and specific shifts are shown for all outcomes in the appendix, as well as coefficient plots with p values constructed using the specific shifts.

For comparison purposes, panel a) repeats the plot for agriculture from the 2001-2011 household survey data. Panel b) likewise examines whether the individual works in agriculture, using data from the 1980 Population Census, again limiting the sample to prime age males. The patterns are similar to those from 2001-2011, but the effects are even larger than in the more recent period. Places in close proximity to a factory are 30 to 35 percentage points less agricultural than those just ten to twenty kilometers away. Indonesia was much more agricultural in 1980, and these results suggest some convergence with more recent industrialization. The appendix shows that the magnitudes remain similar when the means from the counterfactual distributions are differenced out.

Next, panels c) and d) consider whether the individual works in manufacturing, again using data from the 2001-2011 household surveys and the 1980 Population Census. The pattern for manufacturing during 2001-2011 is the inverse of what we see for agriculture, but noisier, with around five percentage points more individuals working in manufacturing. The 1980 Census likewise reveals that manufacturing employment is higher within a few kilometers of the historical factory, though again effects are noisy. Places near a historical factory had around seven percentage points more individuals working in manufacturing, a very large effect given that Indonesia was a primarily agrarian economy at the time – only 11 percent of the population in 1980 was employed in manufacturing. The effects remain unchanged when the counterfactual means are differenced out, indicating that it is not simply the case that places near suitable factory locations are richer.

We also examine employment in retail. For the 2001-2011 period, places within the immediate vicinity of the factory have around 10 percentage points more employment in retail relative to places ten to twenty kilometers away, and the coefficient for the closest bin is above the 90th percentile of the counterfactual distribution. In contrast to manufacturing,

the retail effects are somewhat higher today than in 1980 in the immediate vicinity of the factory.

The occupational patterns – more manufacturing, more retail, and less agricultural – are consistent with the areas near historical factories being more urban. Moreover, historically the sugar factories were major centers of employment, which increasingly relied on wage labor as time went by, and population would have plausibly clustered nearby. To explore this, panels g and h examine log population density, using data from 2003 and 1980 (respectively), defined as the log of population in each *desa/kelurahan* per square kilometer. Areas within 1km or so of a factory are much more densely populated than area 10 km away in both 1980 and today.

4.1.2 Sugar and Linked Industries

There are a number of reasons why colonial sugar factories might influence economic activity in the long run. Most obviously, while the technology has not been water-powered for over a century and sugar cane is no longer central to Indonesian agriculture, the sugar factories themselves could have persisted.

Figure 7, panel a) shows using data from the 2006 Economic Census that indeed, places near a historical factory - particularly those within 0 to 1 kilometers - produce substantially more processed sugar today than places further away. In contrast, panel b) shows that once we drop historical factories within 2 kilometers of a modern factory from the analysis, which represent only 19% of historical factories, the relationship between distance to the factory and modern sugar production flattens out.

Panel c) shows that there is little relationship between raw cane production, drawn from Podes 2003, and distance to a historical factory, presumably because improvements in transport have obviated the need to grow cane in the immediate vicinity of processing plants. The relationship between modern sugar cultivation and distance to a historical factory remains flat in this restricted sample (panel d).

Input-output linkages were plausibly an important driver of agglomeration around sugar factories historically, since low grade sugar was costly to transport. To test whether such agglomeration has persisted, we construct weighted average employment shares for industries upstream and downstream from sugar processing. Sectoral employment is drawn from the 2006 Economic Census and total employment is from the 2000 Census. The weights are from the Leontief inverse of the 2006 Indonesian Input-Output Table. The input-output table specifies how many dollars of sector i 's output are needed by sector j to produce one dollar of its own output. The higher this number, the stronger the linkages between the sectors. The Leontief inverse of this matrix captures not just direct linkages but also indirect

ones. If sector k uses inputs from sector j , which in turn uses inputs from sector i , sector k is indirectly linked to sector i via sector j . The Leontief inverse measures how much of sector i 's output is used both directly and indirectly by sector k . The Leontief weights are described in more detail in the data appendix.

The main manufacturing sectors upstream from sugar processing are farm machinery, used to harvest cane, and capital equipment, used to process cane. (Raw cane is, of course, the largest upstream sector but is not included in our measure since the data only include manufacturing establishments.) The main manufacturing industries downstream are in food processing, as sugar is an additive to most processed foods. Many types of services - restaurants, hotels, schools, and hospitals, to name a few - are also downstream from sugar processing, since sugar is consumed at these establishments, but these are not included in the downstream measure since they are not in manufacturing. These downstream linkages could though contribute to the retail impacts documented above.

Figure 7, panel e) shows that when all historical factories are included in the analysis, upstream industries - mostly farm and processing machinery - are about three times as prevalent relative to the mean within one kilometer of historical factories. This relationship flattens out when we exclude historical factories that are located near modern ones (panel f).

In contrast, employment in manufacturing industries downstream from sugar is much higher near the historical factories, even when we limit the sample to historical factories that are not near modern ones (panels g and h). This is particularly true for places within 0 to 1 kilometer of a historical factory, where the employment share in downstream industries is about 50 percent higher than the mean in villages ten to twenty kilometers away. Sugar was used as an input in other foods historically, and other processed foods also tend to be used as inputs into each other. This suggests a particular channel for manufacturing persistence: even after the original sugar factories disappeared, there were still agglomeration advantages for the remaining downstream firms to continue to locate in the same place, and these persisted over time.

4.1.3 Public goods and the public sector

The analysis thus far has focused on the private sector, but there may be important linkages that occurred through the location of public investments. The historical literature emphasizes that the Dutch government constructed road and rail networks to transport sugar to ports. The configuration of transport infrastructure tends to be very persistent, and hence these impacts may have well persisted through the present.

Figure 8, panel a) documents that intercity road density is higher today in the immedi-

ate vicinity of factories, and these effects are atypical of those near counterfactual factories. Moreover, data from the 1980 Podes, a census of village governments, similarly show that villages in the immediate vicinity of the historical factories were less likely to only be accessible via a dirt road historically (panel b).

The 1980 Podes data also reveal that places within a few kilometers of the factories were substantially more likely to have electricity, and the effects fall far in the right tails of the counterfactual distributions. In the immediate vicinity of the historical factories, villages were 50 percentage points more likely to have electricity than places ten to twenty kilometers away. In 1980 relatively few households had electricity, and hence higher rates of electrification likely indicate a greater presence of manufacturing.

In 1980, areas in the immediate vicinity of a factory were also more likely to have high schools, which were typically in centrally located towns (panel d). Indeed, places near factories are around two kilometers closer to the nearest subdistrict capital, relative to a sample mean distance of 3.8 kilometers (panel f). The effects on high schools are also positive using the pooled 1996-2011 Podes Village Censuses but are not statistically significant relative to the placebo distributions, and high schools had become more prevalent by this period (panel e).

Figure 9 examines long-run impacts on schooling in more detail, using cohort level data from the 2000 Population Census. We focus in the main text on three representative cohorts, with the remainder shown in the appendix: the 1920-1929 cohort (educated during the Dutch period), the 1950-1954 cohort (educated following independence), and the 1970-1974 cohort (educated during Indonesia’s large-scale school building campaign). Panel a) documents that individuals within a few kilometers of a historical factory on average have between 0.75 and 1.25 more years of schooling than those located ten to twenty kilometers away. These effects are atypical relative to the counterfactual distributions and hold across all three cohorts (panel b). We see a similar pattern for primary completion (panels c and d). It is much higher in close proximity to the historical factories, and this is particularly true for the two older cohorts, whose schooling occurred at a time when primary access was far from universal. High school completion again shows a similar pattern (panels e and f). Effects on high school are largest for the younger cohorts, who received schooling at a time when high school was more common, but impacts are still positive and statistically significant even for the oldest cohort, educated during the Dutch era.

4.1.4 Household Consumption

While we cannot separately identify whether improved infrastructure comes from these areas being manufacturing centers or administrative centers (or both), there is clear evidence that

they are more developed in a variety of ways. We conclude by examining the effects on household consumption, using pooled data from the SUSENAS. Following Deaton (1997), we assume that children aged 0 to 4 are equal to 0.4 adults and children aged 5 to 14 are equal to 0.5 adults. All regressions control for survey year fixed effects and the number of household members aged 0-4, 5-14, and 15 and older. We find that consumption levels in areas immediately adjacent to the Cultivation System factories is around 14 percent higher than areas even just 6 kilometers further away. Consistent with these being relatively integrated areas, this effect is consistent with we would expect given the differences in education we observed before – people living adjacent to a Cultivation System factory have about 1.4 more years of education, so an 8-10 percent return to schooling (Duflo, 2001) would yield exactly the consumption differences we observe.

4.2 Discussion

In summary, the Dutch colonial sugar industry substantially transformed economic activity in contemporary Java. Sectoral composition suggests that industries that used sugar as an input sprung up surrounding colonial sugar factories, creating food processing centers with population clustered nearby. Even when the original source of sugar disappeared, those industrial centers persisted, sourcing sugar from elsewhere. Moreover, historically infrastructure was required to transport sugar to ports, and we show that there are still impacts on public infrastructure today. These results highlight how colonial extraction can have a range of effects, some of which may encourage agglomeration or infrastructure investments that promote economic activity in the long run.

5 Impacts on Subjected Villages

5.1 Empirical Strategy

The discussion thus far has focused on one aspect of the Cultivation System, namely the establishment of the factory itself. The second part of our analysis turns to the impact on the villages that were forced to grow sugar cane.

To estimate the effects of forced cultivation, we exploit the discontinuous change in exposure at the borders of the subjected catchment areas (as illustrated by Figure 1). Inside this catchment area, villages cultivated sugar for the government, whereas outside they did not. The boundaries form a multi-dimensional discontinuity in longitude-latitude space, and

regressions take the form:

$$out_v = \alpha + \gamma sugar_v + f(\text{geographic location}_v) + g(dfact_v) + \beta \mathbf{X}_v + \sum_{i=1}^n seg_v^i + \epsilon_v \quad (2)$$

where out_v is the outcome variable of interest in village v . $sugar_v$ is an indicator equal to 1 if the modern village is matched with a historical village that grew sugar cane for the Cultivation System and equal to zero otherwise. $f(\text{geographic location}_v)$ is the RD polynomial, which controls for smooth functions of geographic location. Following Gelman and Imbens (2014), we use a local linear RD polynomial for the baseline and document robustness to a wide variety of different bandwidths and RD polynomials.¹⁴ The RD polynomial is estimated separately for each of the 17 contiguous catchment areas. To ensure that the effects we estimate are due to a village being subjected to growing cane, rather than merely being close to a factory, $g(dfact_v)$ controls for a linear spline in distance to the nearest historical sugar factory, with kink points estimated every three kilometers; Section 4 shows that this captures the impacts of proximity to a factory well. $\mathbf{X}_v$ contains exogenous geographic characteristics: elevation, slope, distance to the coast, distance to the closest natural harbor, distance to the nearest river, log flow accumulation, and distance to the nearest 1830 residency capital. The seg_v^i split each catchment area boundary into 10 kilometer segments, equaling one if village v is closest to segment i and zero otherwise. They ensure that the specification is comparing nearby villages; the appendix shows that results are robust to the choice of segment length. For regressions examining household consumption, we also include a vector of demographic variables giving the number of infants, children, and adults in the household. The baseline specification limits the sample to villages within 10 kilometers of the threshold.¹⁵ Standard errors are clustered at the sub-district level.

Villages where sugar could not grow are not a suitable counterfactual for areas with government cultivation. Hence, we limit the sample to areas that are suitable for sugar, where suitability is defined using the same elevation and slope requirements described in Section 4. For the same reason, we exclude places in cities that were residency capitals.

The key identifying assumption is that all relevant factors besides treatment vary smoothly at the Cultivation System boundaries. That is, letting c_1 and c_0 denote potential outcomes under treatment and control, x denote longitude, and y denote latitude, identification requires that $E[c_1|x, y]$ and $E[c_0|x, y]$ are continuous at the discontinuity thresholds. This assumption is needed for observations located just outside the catchment areas to be an appropriate counterfactual for observations located just inside.

¹⁴Regressions use a triangular kernel such that the weight given to each observation decays with distance from the threshold.

¹⁵Treated villages are never more than ten kilometers from the nearest catchment boundary.

To assess the plausibility of this assumption, Table 1 examines a variety of geographic characteristics, using gridded geographic data and regressions of the form described in equation (2). RD graphs for these and all subsequent RD results in this section are shown in the appendix. Column (1) examines elevation. The point estimate on *cultivation village* is negative and statistically significant, but the difference is only 2 meters, which is much too small to have an appreciable effect on weather or agricultural potential. Indeed, the entire sample is in the plains, very close to sea level (the mean elevation in the sample is 31 meters). None of the remaining characteristics – slope, flow accumulation, a dummy for being on the coast, distance to the coast, nearest river, or nearest natural harbor, or distance to the 1830 residency capital – show any statistically significant or economically meaningful differences.¹⁶

Another RD identifying assumption is that individuals did not selectively sort around the threshold while the Cultivation System was in force, in order to exploit its design. Typically in this context, one would worry that a substantial number of productive individuals moved just outside the subjected areas to escape forced cultivation. However, as will be shown below, we find positive economic impacts of the Cultivation System on subjected villages, and it appears implausible that high productivity individuals would have moved to regions subjected to forced cultivation. In historical Java, individuals who migrated to an already established village were not eligible to hold land, and disempowered movers would have plausibly borne the brunt of forced labor.¹⁷ Alternatively, low productivity individuals may have fled subjected villages.¹⁸ However, population density today is if anything greater in treated villages, suggesting that mass out-migration is unlikely to drive results.

5.2 Results

5.2.1 Land

The historical literature has largely focused on how the Cultivation System impacted land tenure in subjected villages. In particular, village heads under the Cultivation System were empowered by the Dutch to redistribute land in order to facilitate the growing of sugar cane. As discussed in Section 2, the literature emphasizes that land was redistributed to the village,

¹⁶One might also be interested in examining other characteristics, such as climate or crop suitability. However differences tend to be zero by construction, since treated and non-treated villages along the boundaries are at most a kilometer or two apart and hence often fall within the same cell due to the lower resolution of these data. Similarly, while ideally we would be able to look at social and economic characteristics before the Cultivation System, the very limited pre-period data are at a much coarser level of aggregation.

¹⁷While in theory landowners were responsible for cultivation services, in practice landless peasants attached themselves to landholding families and often performed the labor.

¹⁸Elson (1984, p. 60) indeed argues that the rapid growth of frontier regions of interior Java during the Cultivation System may have been due to individuals fleeing forced labor, though it is not obvious that the lowest productivity individuals would have been the ones who moved.

whether to directly benefit officeholders or to facilitate sharing the burdens of the System (Elson, 1984, p. 94; Ricklefs, 2008, p. 159). Village land is used for multiple purposes, but an important one is as *tanah bengkok*, which is land owned by the village which the village head is allowed to use (or rent out) as compensation for his service.

We therefore begin by examining the allocation of land in more contemporary periods, and in particular focusing on village-owned land, as well as overall land inequality among privately held land. Indonesian PODES village censuses from 2003 and 1980 collected information on village-owned land, and land inequality for private agricultural land can be measured using the 100% microdata sample from the 2006 Agricultural Census.

The estimates show that the Cultivation System left a substantial mark on the amount of village-owned land available in contemporary villages. Table 2, columns (1) through (4) show that in both 2003 and 1980, Cultivation System villages had substantially more village-owned land, in absolute terms (columns 1 and 3) and as a percentage of total land (columns 2 and 4). The differences are highly statistically significant. In 2003, about 1.4 percentage points more land was owned by the village in Cultivation System areas, relative to a sample mean of 9% of total land being owned by villages. In 1980 these villages owned 1.2% more land, relative to a sample mean of 11%. RD plots are shown in the appendix.

Columns (5) through (8) examine land inequality amongst agricultural households, considering the 99-90 ratio, the 90-10 ratio, the 90-50 ratio, and the 50-10 ratio of agricultural land usage.¹⁹ All effects are positive, but they fall short of statistical significance.

5.2.2 Public goods

The existing literature suggests that the village's ability to raise revenue, as well as land tenure arrangements more generally, could impact human capital accumulation and the provision of local schooling (Dell et al., 2017; Acemoglu et al., 2009; Banerjee and Somanathan, 2007). This may have been particularly true in the Dutch period and in independent Indonesia prior to the massive INPRES school expansion, studied by Duflo (2001), as prior to INPRES most village-level school construction was financed locally by the villages themselves (Aritonang, 1994).

We begin by looking at the presence of schools in 1980, the earliest date for which systematic village level data on different types of schools are available. Table 3 examines the presence of schools using data from the Podes 1980 village census. 1980 was around the time of Indonesia's massive primary school building campaign - INPRES - which largely equalized access to basic primary schooling. The data break down primary schools into those that were built by the INPRES campaign and those that were not. Columns (1) and (2) focus on non-

¹⁹Ownership data do not distinguish between agricultural and non-agricultural land, and hence we unfortunately cannot examine inequality in land ownership.

INPRES schools. There is not an impact on the number of buildings, whereas the effect on the number of teachers is if anything positive but falls just short of statistical significance. Columns (3) and (4) show that these villages received significantly fewer INPRES schools and teachers, almost surely an indicator of greater education beforehand since INPRES was targeted based on the availability of pre-existing primary education (Duflo, 2001). Subjected villages were around four percentage points less likely to receive an INPRES school, relative to a sample mean of 0.36. Column (5) shows that subjected villages were 2 percentage points more likely to have a junior high school, relative to a sample mean of only 6 percent of villages that had them (junior high schools were not covered by INPRES). There is not a statistically significant impact on the presence of a high school, which only two percent of villages had.

These impacts are consistent with the fact that during the Dutch period, if communities wanted a school they were responsible for raising a share of the required resources. In particular, under the *volkschool* (village school) program, the colonial government would provide resources towards hiring a teacher only if the community constructed and maintained the school (Aritonang, 1994).

We then look at whether there are persistent effects on educational attainment, using the 100% sample of the 2000 Population Census. Table 4, column (1) documents that individuals in subjected villages have around 0.22 years more schooling, relative to a sample mean of 5 years, and the effect is statistically significant at the 1% level. They are also more likely to complete primary school and junior high (columns 2 and 3). There is not a statistically significant impact on high school completion, though only 13% of the sample completed high school (column 4).

Columns (5) and (6) examine the 1980 Population Census. Individuals in subjected villages are 1.7 percentage points less likely to have no schooling, relative to a sample mean of 41%. There is not an effect on primary completion, with only 19% of the sample in 1980 completing primary school.

To examine impacts on education even further back in time, we estimate effects by cohort. For the 1980 Census, data are only available for a relatively small sample, and hence cohort estimates are quite noisy. In contrast, the 2000 Census provides a 100% sample, so we can estimate village-level impacts on each cohort, beginning with the cohort born between 1920 and 1930 and continuing through the cohort born between 1975 and 1980, the youngest cohort to have reached adulthood by 2000.²⁰ The left panel in Figure 11 plots estimates from equation (2) by cohort, so each point in each panel in the left-hand side of Figure 11 represents a γ coefficient from a separate estimation of equation (2), with the 95% confidence interval indicated by a line. For ease of interpretation, the right panel plots sample means

²⁰We use ten year age groupings to increase power for the oldest cohorts, since these are much smaller.

for each cohort and outcome variable. In general, schooling levels were initially very low, and increase over time.

The left panel of Figure 11 shows that impacts on years of schooling are large and positive across cohorts, even going as far back as the cohort born in the 1920s who would have completed their primary education under the Dutch. Effects on primary completion peak for cohorts born when primary was rapidly expanding - in particular for cohorts born in the 1950s and 60s. The impact decreases somewhat as primary completion becomes more universal, though is still present in the most recent cohorts. On the other hand, impacts on junior high completion are zero for the older cohorts, whose junior high completion rates were close to zero, and then become large and positive as cohorts born from the 1950s onward begin to complete junior high at more appreciable rates. Impacts on high school completion show a similar pattern.

Finally, columns (7) and (8) of Table 4 consider the education of village heads, pooling data from Podes village censuses collected between 1996 and 2011. The regressions include survey year fixed effects. While there is not a statistically significant impact on years of schooling, village heads in subjected areas are three percentage points more likely to have completed high school, relative to a sample mean of 74%. The effect on village heads is about double the average effect for cohorts of similar age shown in Figure 11, panel g). One possibility is that the more generous compensation for village heads, in the form of more village land they are allowed to use ex-officio, attracted a relatively more educated pool of village heads.

5.2.3 Economic structure

We finally turn to the overall structure of the economy. Table 5, columns (1) through (3) examine data from the SUSENAS household survey, collected in annual waves from 2001 through 2011. All regressions include survey year fixed effects.

We find that villages subjected to the Cultivation System are substantially *less* agricultural today. Specifically, in subjected villages, individuals are 4 percentage points (15 percent) less likely to work in agriculture, 2.9 percentage points (14 percent) more likely to work in manufacturing, and 1.2 percentage points (7 percent) more likely to work in retail. The sample is limited to prime aged males - aged 18 to 55 - to avoid confounding impacts with labor force participation, but estimates are similar when the entire labor force is included.

Columns (4) through (6) examine the structure of the economy in 1980, using data from the 1980 Population Census and again focusing on prime age males.²¹ The coefficient on

²¹We do not consider sectoral information from the 2000 Census, as it was collected in the midst of a severe economic downturn – the Asian Financial Crisis – and hence is less informative about long-run economic

employment in agriculture is similar in magnitude to that from more modern SUSENAS data, but noisily estimated. The effect on manufacturing, which was a much smaller share of the Indonesian economy in 1980, is close to zero and statistically insignificant. However, we cannot rule out a similar proportionate increase in manufacturing to that in the more recent data. Individuals in subjected villages were 2.7 percentage points (23 percent) more likely to be employed in commerce.

Data from Indonesia’s 2006 Economic Census, examined in Table 6, show a similar pattern. The Economic Census captures all large manufacturing firms and a random sample of small and informal firms. We find that subjected villages have 30 percent more manufacturing firms than non-subjected villages (column 1). Column (2) considers the village’s manufacturing employment, taken from the Economic Census, divided by the size of the labor force, drawn from the 2000 Population Census. The estimate, which indicates a 2.4 percentage point (15 percent) increase, is similar to what we obtain from the SUSENAS household surveys, though more noisily estimated (not surprisingly since it is drawn from multiple datasets). Population density is also around seven percent higher in subjected villages, both in 2003 and in 1980.

Finally, column (5) examines equivalent consumption from the SUSENAS household survey. There is not a statistically significant difference in household consumption. However, the confidence intervals include the possibility of consumption being around 2 percent higher, about what we would expect given that individuals in subjected villages on average have around 0.2 years additional schooling.

5.2.4 Discussion

If anything, the long-run development effects of forced cultivation reinforce the impacts of the sugar factories documented above. This result is somewhat surprising in the context of other studies that find negative long-run development effects of forced colonial labor (Lowes and Montero, 2016; Dell, 2010), but the historical and empirical evidence offer various clues about what drives the impacts on cultivation villages.

The historical literature emphasizes that while in force, the Cultivation System led to increases in communal village landholdings and the share of village lands apportioned to village officeholders (Elson, 1984, p. 94; Ricklefs, 2008, p. 159). We find that still today there is about 10 to 15 percent more public-use land in Cultivation System villages. This has remained fairly constant over time, with very similar effects in both 1980 and 2003. Households in Cultivation System villages are also more educated going all the way back to the 1920s cohort, which received its education during the Dutch period. While there

conditions.

are a variety of channels that could link the Cultivation System to long-run human capital accumulation, the presence of village land is a particularly plausible mechanism. During the early 20th century, villages that wanted a school needed to fund the school building themselves, and the revenue from village-owned lands may have facilitated this construction.

The historical literature also emphasizes that the influx of crop incentive payments into Cultivation System villages incentivized modern production and exchange (Elson, 1994, p. 261, Ricklefs, 2008, p. 158) Consistent with this, we find that cultivation villages continue to have a different economic structure, with a greater percentage of households working in manufacturing and retail and fewer in agriculture, both in 1980 and today. Higher human capital levels likewise might lead to less agricultural production.

6 Conclusion

The economic changes induced by the Cultivation System led to a more modern economy, with more manufacturing and retail and greater education levels. This is true both for the areas in the immediate vicinity of the historical sugar factories and in the villages where the economy was reorganized to grow sugar cane. Input-output linkages, agglomeration, infrastructure investments, and human capital accumulation are plausible channels of persistence. Colonial extraction took many forms and had diverse effects, including directly on the structure of the economy. These results plausibly inform a variety of contexts in which colonizers attempted to extract surplus by reorganizing the economy along more modern lines.

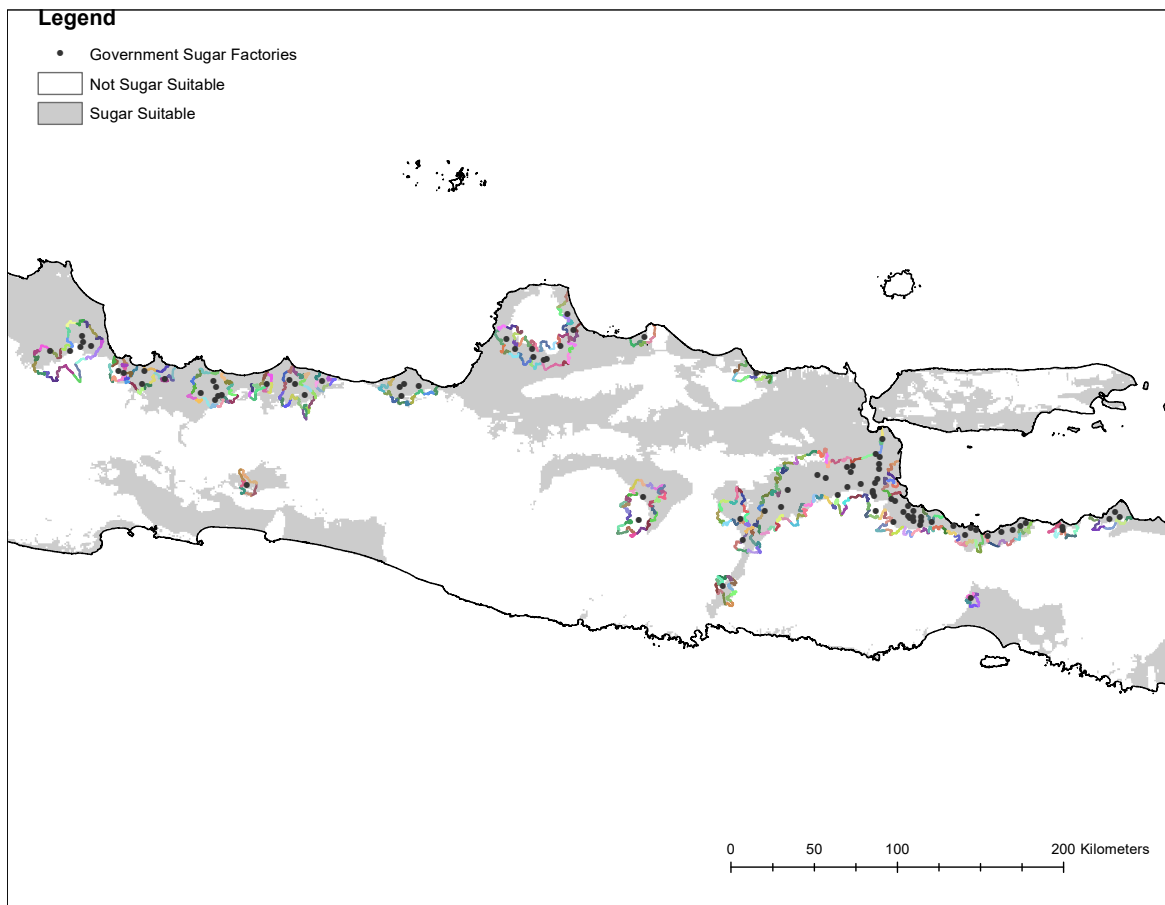
References

- ACEMOGLU, D., U. AKCIGIT, AND W. KERR (2016): “Networks and the Macroeconomy: An Empirical Exploration,” *NBER Macroeconomics Annual*, 30, 273–335.
- ACEMOGLU, D., M. A. BAUTISTA, P. QUERUBIN, AND J. A. ROBINSON (2009): “Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia,” in *Institutions and Economic Performance*, ed. by E. Helpman, Harvard University Press.
- ACEMOGLU, D., D. CANTONI, S. JOHNSON, AND J. A. ROBINSON (2011): “The Consequences of Radical Reform: The French Revolution,” *American Economic Review*, 101, 3286–3307.
- ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): “The Network Origins of Aggregate Fluctuations,” *Econometrica*, 80, 1977–2016.
- ACEMOGLU, D., C. GARCIA-JIMENO, AND J. A. ROBINSON (2015): “State Capacity and Economic Development: A Network Approach,” *American Economic Review*, 105, 2364–2409.
- ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): “The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review*, 91, 1369–1401.
- (2002): “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution,” *The Quarterly Journal of Economics*, 117, 1231–1294.
- ACEMOGLU, D. AND J. A. ROBINSON (2013): *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, Crown Business.
- ALESINA, A., P. GIULIANO, AND N. NUNN (2013): “On the Origins of Gender Roles: Women and the Plough,” *The Quarterly Journal of Economics*, 128, 469–530.
- ARITONANG, J. S. (1994): *Mission Schools in Batakland (Indonesia): 1861-1940*, vol. 10, Brill.
- BANERJEE, A. AND L. IYER (2005): “History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India,” *American Economic Review*, 95, 1190–1213.
- BANERJEE, A. AND R. SOMANATHAN (2007): “The Political Economy of Public Goods: Some Evidence from India,” *Journal of Development Economics*, 82, 287–314.
- BAQAEI, D. (2015): “Labor Intensity in an Interconnected Economy,” Working paper, Harvard.
- BECKER, S. O., K. BOECKH, C. HAINZ, AND L. WOESSMANN (2016): “The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy,” *The Economic Journal*, 126, 40–74.
- BUKOWSKI, P. (2016): “How History Matters for Student Performance. Lessons from the Partitions of Poland,” Working paper, Central European University, http://www.personal.ceu.hu/staff/repec/pdf/2015_3.pdf.
- CARVALHO, V. M. (2009): “Aggregate Fluctuations and the Network Structure of Intersectoral Trade,” Working paper, Pompeu Fabra University.
- COMMISSIE UMBGROVE (1858): “Archief van de Commissie voor de Opname van de Verschillende Suikerfabrieken op Java [1853–1858],” Nationaal Archief, The Hague, Access Number 2.10.11.
- DEATON, A. (1997): *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, World Bank.
- DELL, M. (2010): “The Persistent Effects of Peru’s Mining Mita,” *Econometrica*, 78, 1863–1903.

- DELL, M., N. LANE, AND P. QUERUBIN (2017): “The Historical State, Local Collective Action, and Economic Development in Vietnam,” Working paper, Harvard.
- DUFLO, E. (2001): “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review*, 91, 795–813.
- ELSON, R. E. (1984): *Javanese Peasants and the Colonial Sugar Industry: Impact and Change in an East Java Residency, 1830–1940*, Oxford University Press.
- (1994): *Village Java Under the Cultivation System, 1830–1870*, Asian Studies Association of Australia.
- FASSEUR, C. (1992): *The Politics of Colonial Exploitation: Java, the Dutch, and the Cultivation System*, Cornell University Southeast Asia Program.
- FERNÁNDEZ, R. AND A. FOGLI (2009): “Culture: An Empirical Investigation of Beliefs, Work, and Fertility,” *American Economic Journal: Macroeconomics*, 1, 146–177.
- GEERTZ, C. (1963): *Agricultural Involution: the Process of Ecological Change in Indonesia*, vol. 11, University of California Press.
- GELMAN, A. AND G. IMBENS (2014): “Why high-order polynomials should not be used in regression discontinuity designs,” Tech. rep., National Bureau of Economic Research.
- GENNAIOLI, N. AND I. RAINER (2007): “The Modern Impact of Precolonial Centralization in Africa,” *Journal of Economic Growth*, 12, 185–234.
- GIULIANO, P. (2007): “Living Arrangements in Western Europe: Does Cultural Origin Matter?” *Journal of the European Economic Association*, 5, 927–952.
- GROSJEAN, P. (2011): “The Weight of History on European Cultural Integration: A Gravity Approach,” *American Economic Review*, 101, 504–508.
- (2014): “A History of Violence: The Culture of Honor and Homicide in the US South,” *Journal of the European Economic Association*, 12, 1285–1316.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2008): “Alfred Marshall Lecture: Social Capital as Good Culture,” *Journal of the European Economic Association*, 6, 295–320.
- (2016): “Long-Term Persistence,” *Journal of the European Economic Association*, 14, 1401–1436.
- HAYAMI, Y. (1980): “Economic Approach to Village Community and Institutions,” *Journal of Rural Development*, 3, 27–49.
- HIRSCHMAN, A. O. (1960): *The Strategy of Economic Development*, Yale University Press.
- LIU, E. (2017): “Industrial Policies and Economic Development,” Tech. rep., MIT.
- LOWES, S. AND E. MONTERO (2016): “Blood Rubber: The Effects of Labor Coercion on Institutions and Culture in the DRC,” Working paper, Harvard, <http://www.saralowes.com/research.html>.
- LOWES, S., N. NUNN, J. A. ROBINSON, AND J. WEIGEL (2015): “The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom,” Working Paper 21798, National Bureau of Economic Research.
- LUITEN VAN ZANDEN, J. (2010): “Colonial State Formation and Patterns of Economic Development in Java, 1800–1913,” *Economic History of Developing Regions*, 25, 155–176.
- LUTTMER, E. F. P. AND M. SINGHAL (2011): “Culture, Context, and the Taste for Redistribution,” *American Economic Journal: Economic Policy*, 3, 157–79.

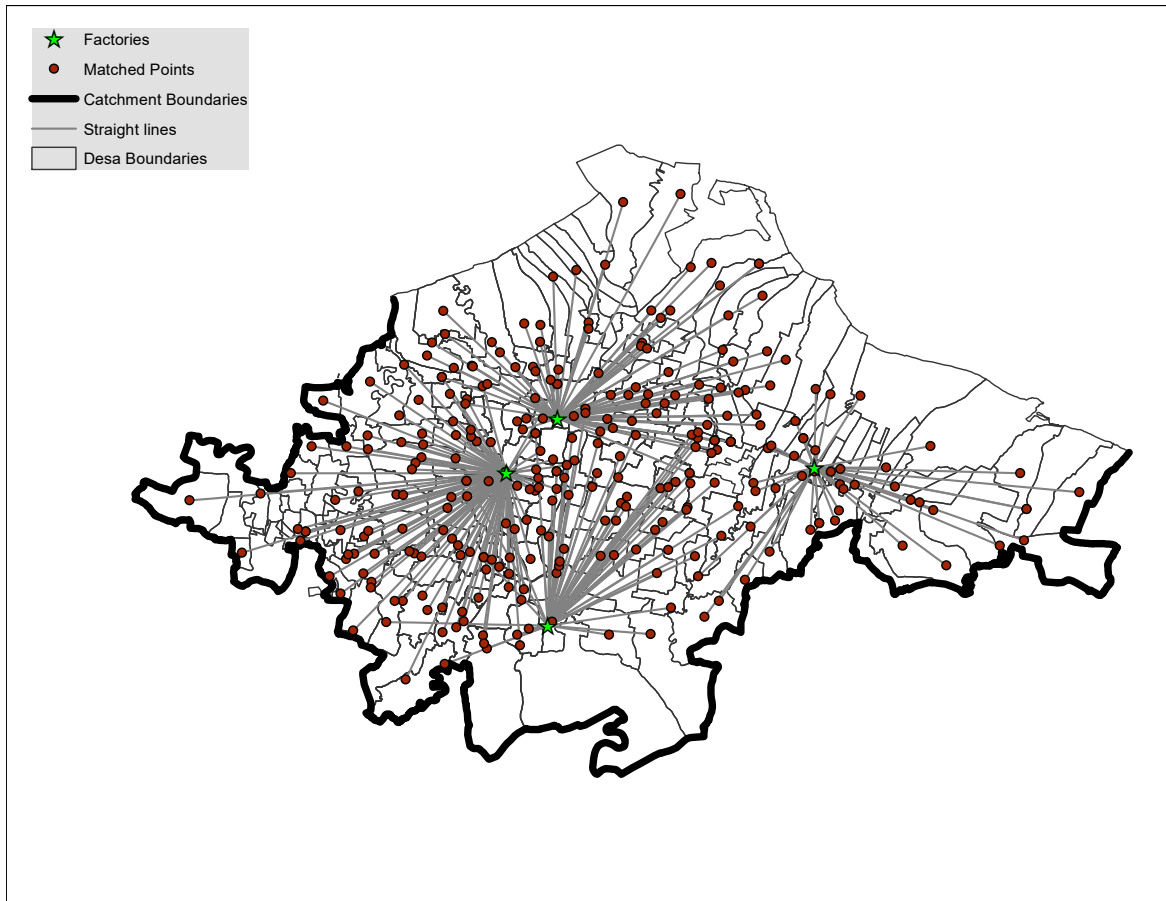
- MICHALOPOULOS, S. AND E. PAPAIOANNOU (2013): “Pre-Colonial Ethnic Institutions and Contemporary African Development,” *Econometrica*, 81, 113–152.
- (2014): “National Institutions and Subnational Development in Africa,” *The Quarterly Journal of Economics*, 129, 151–213.
- MYRDAL, G. (1957): *Economic Theory and Under-Developed Regions*, G. Duckworth.
- NUNN, N. (2008): “The Long-Term Effects of Africa’s Slave Trades,” *The Quarterly Journal of Economics*, 123, 139–176.
- NUNN, N. AND L. WANTCHEKON (2011): “The Slave Trade and the Origins of Mistrust in Africa,” *American Economic Review*, 101, 3221–3252.
- OTO-PERALÍAS, D. AND D. ROMERO-ÁVILA (2014): “The Distribution of Legal Traditions Around the World: A Contribution to the Legal-Origins Theory,” *Journal of Law and Economics*, 57, 561–628.
- RASMUSSEN, P. N. (1956): *Studies in Inter-Sectoral Relations*, E. Harck.
- RICKLEFS, M. C. (2008): *A History of Modern Indonesia Since c. 1200*, Penerbit Serambi.
- SALOP, S. C. (1979): “Monopolistic Competition with Outside Goods,” *The Bell Journal of Economics*, 141–156.
- SCOTT, J. C. (1977): *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*, Yale University Press.
- SPOLAORE, E. AND R. WACZIARG (2013): “How Deep Are the Roots of Economic Development?” *Journal of Economic Literature*, 51, 325–369.
- TABELLINI, G. (2008): “Presidential Address: Institutions and Culture,” *Journal of the European Economic Association*, 6, 255–294.
- (2010): “Culture and Institutions: Economic Development in the Regions of Europe,” *Journal of the European Economic Association*, 8, 677–716.
- TOMLINSON, J. (2014): *Dundee and the Empire: ‘Juteopolis’ 1850-1939*, Edinburgh University Press.
- VAN NIEL, R. (1992): *Java Under the Cultivation System*, KITLV Press.
- (2005): *Java’s Northeast Coast, 1740–1840: A Study in Colonial Encroachment and Dominance*, CNWS Publications.
- VAN SCHAIK, A. (1986): *Colonial Control and Peasant Resources in Java: Agricultural Involution Reconsidered*, Amsterdam: Instituut Voor Sociale Geografie.
- VAN WAIJENBURG, M. (2015): “Financing the Colonial State: The Revenue Imperative and Forced Labor,” Working paper, African Economic History Network.
- VOIGTLÄNDER, N. AND H.-J. VOTH (2012): “Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany,” *The Quarterly Journal of Economics*, 127, 1339–1392.

Figure 1: The Cultivation System



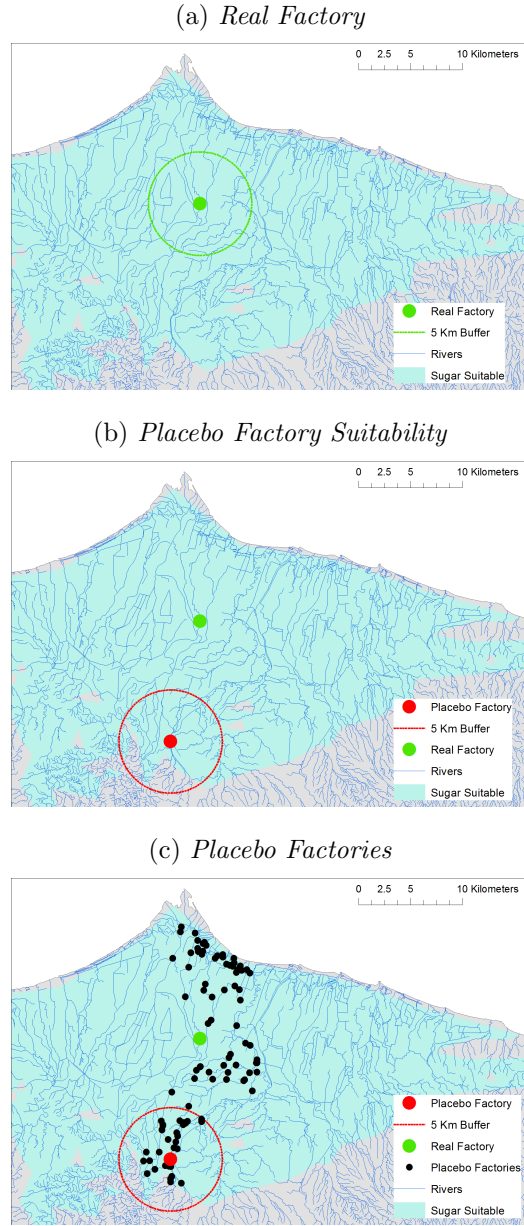
Sources: Commissie Umbgrove (1858).

Figure 2: Construction of Catchment Areas



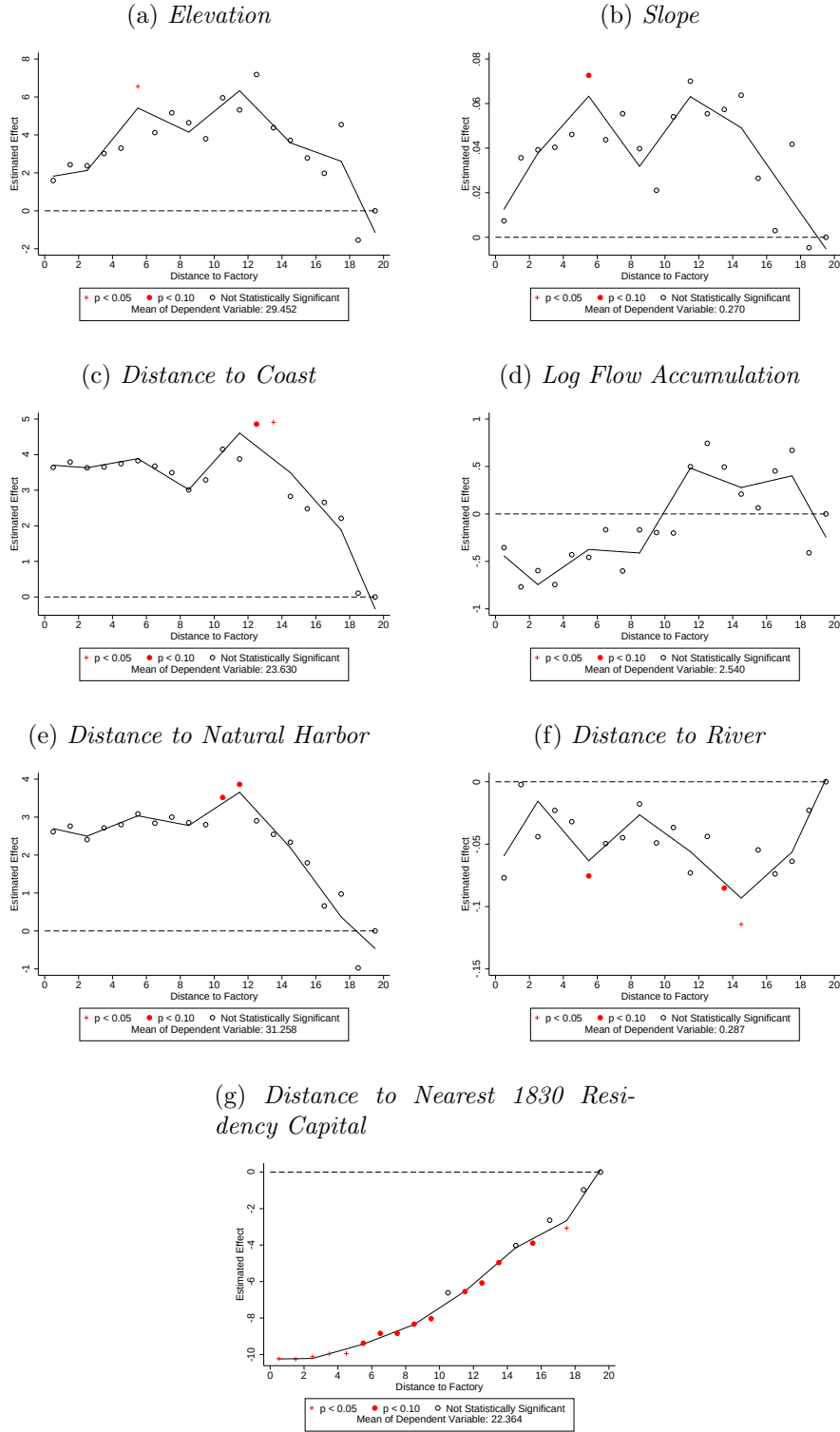
This figure illustrates the construction of the catchment areas, as described in Section 3.1.

Figure 3: Placebo Factories Example



This figure illustrates the construction of the placebo factories, as described in Section 4.1.

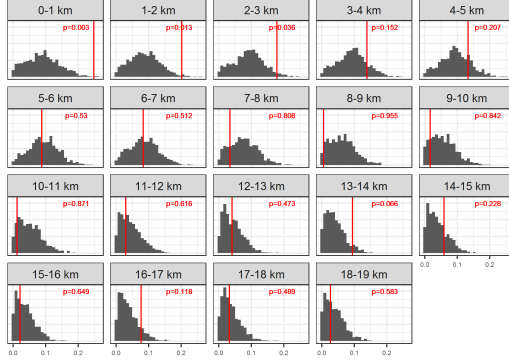
Figure 4: Geography



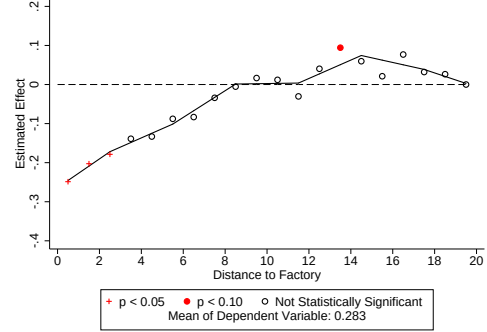
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure 5: Share in Agriculture (2001-11): Illustration of Methodology

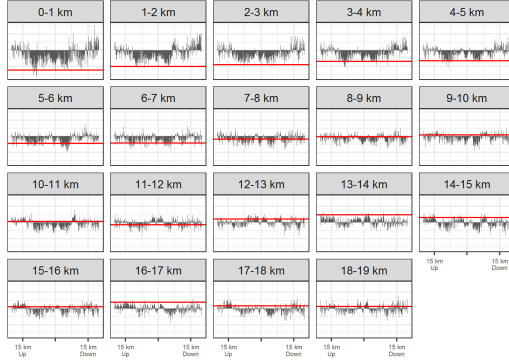
(a) *Independent Shifts: Counterfactuals*



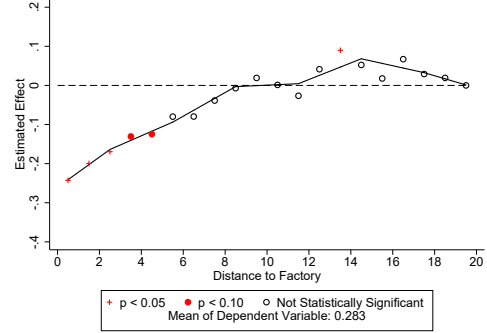
(b) *Independent Shifts: Plotted Coefficients*



(c) *Common Shifts: Counterfactuals*

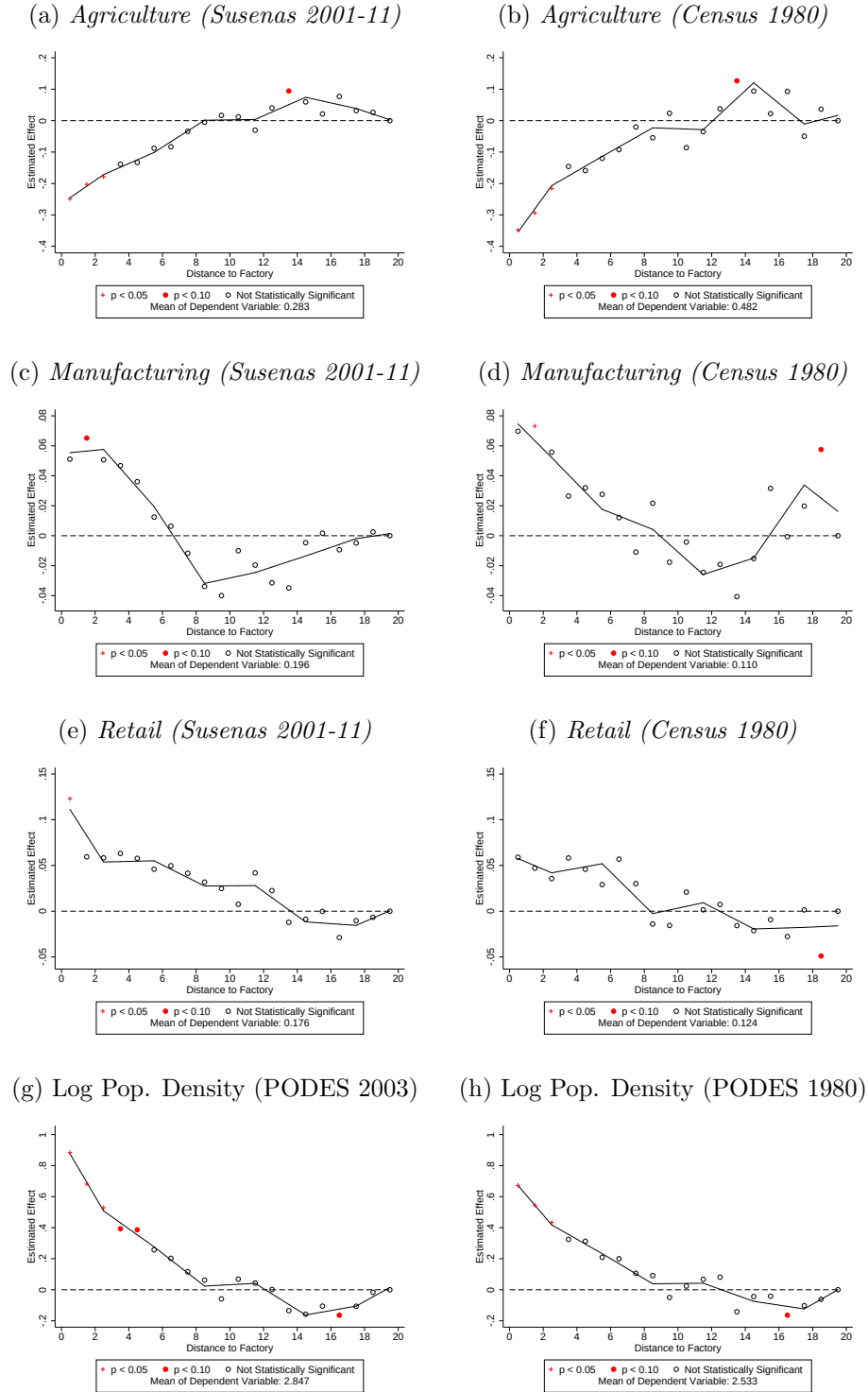


(d) *Common Shifts: Plotted Coefficients*



Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines.

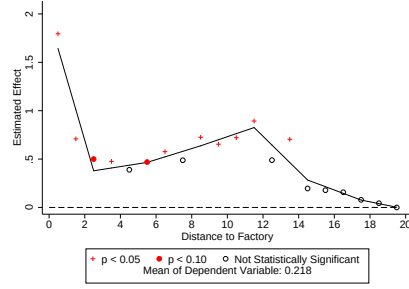
Figure 6: Industry and Agglomeration



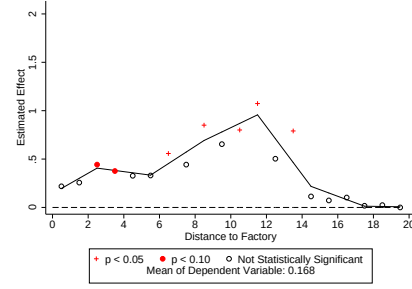
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure 7: Sugar and Linked Industries

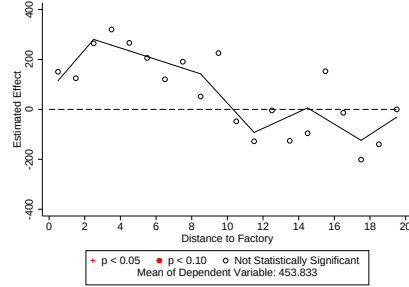
(a) *Log Value Sugar Processed (Full Sample, Economic Census 2006)*



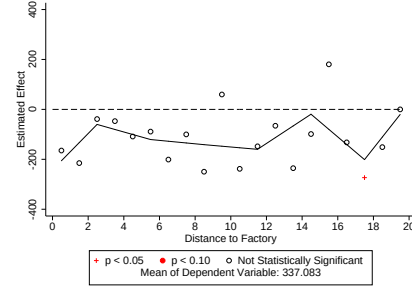
(b) *Log Value Sugar Processed (No Modern Factories, Econ Census 2006)*



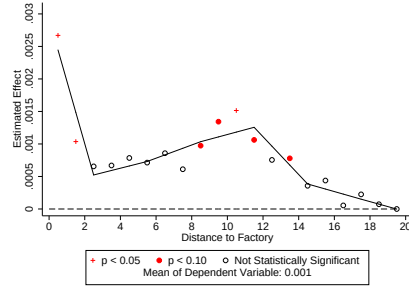
(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



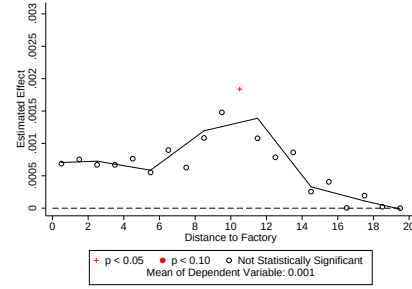
(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



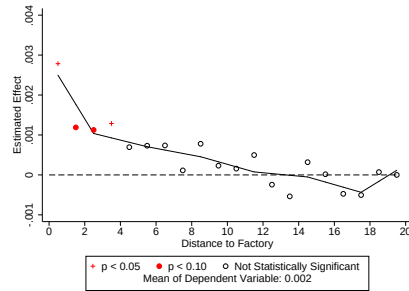
(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



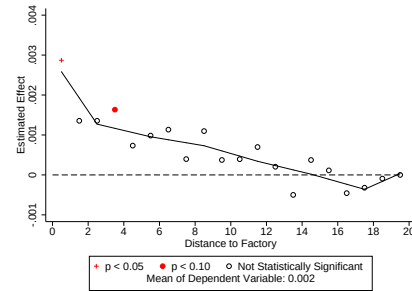
(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*



(g) *Employment Share Downstream (Full Sample, Economic Census 2006)*

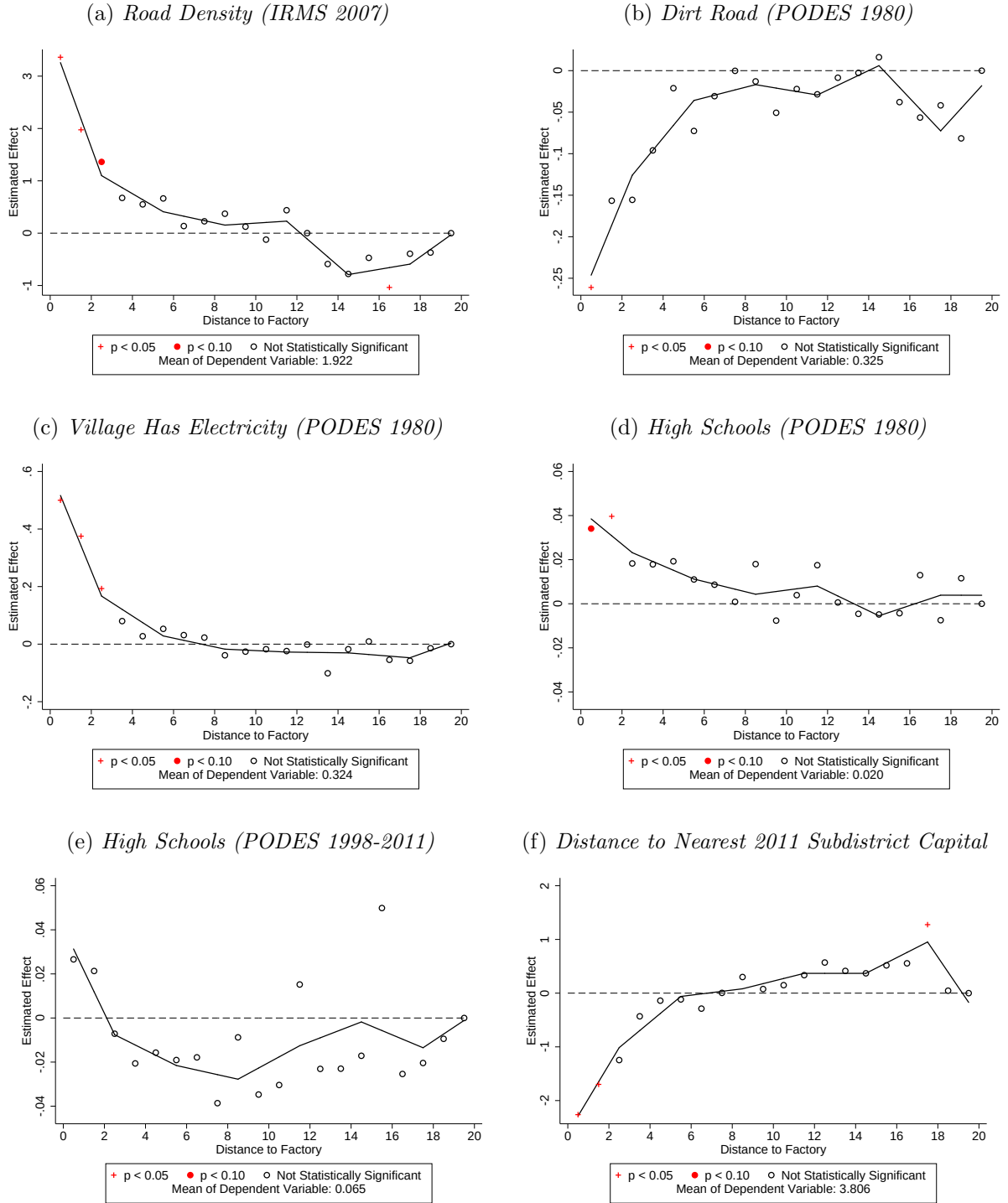


(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



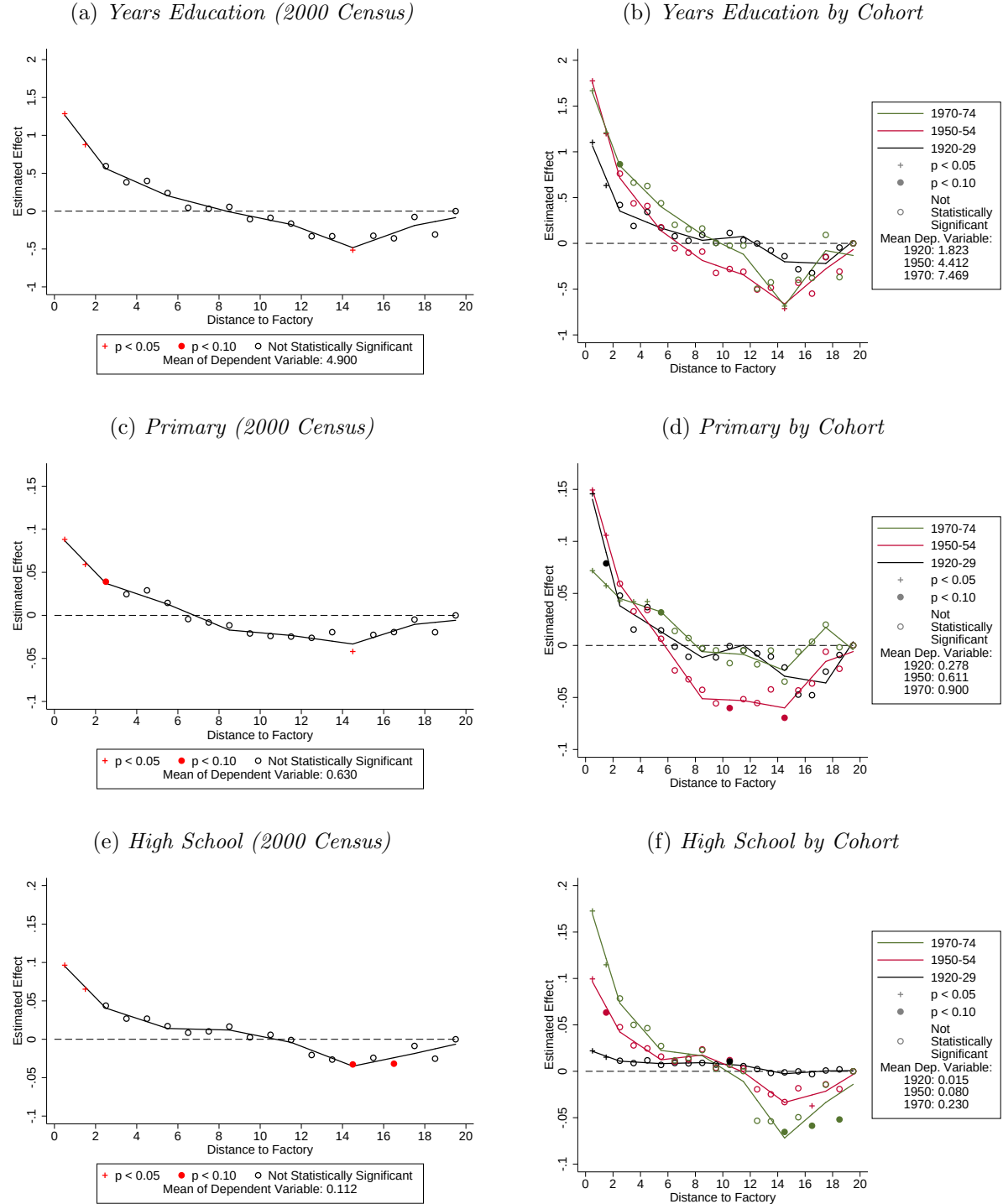
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure 8: Public Goods



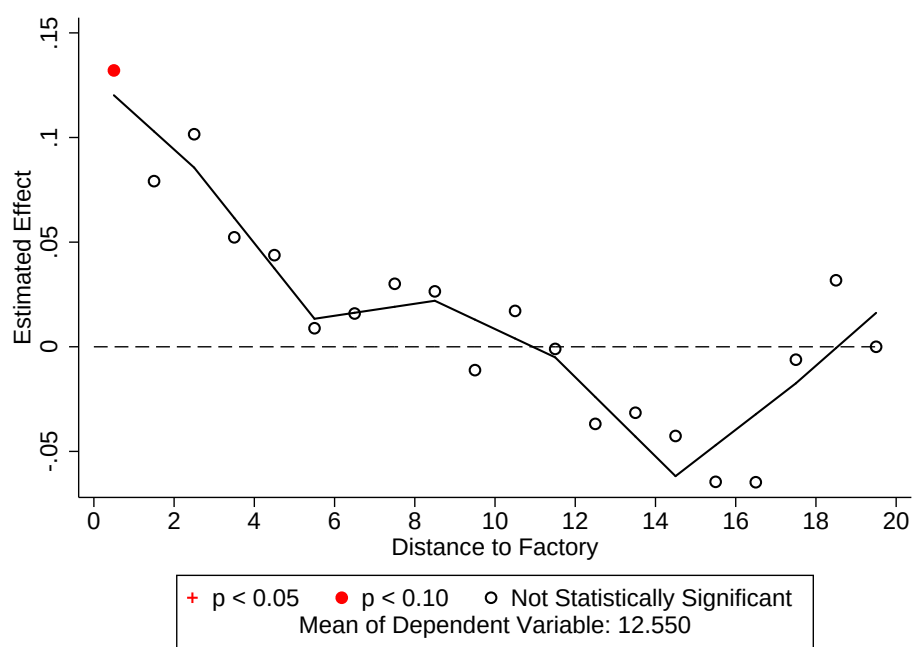
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel e) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure 9: Education



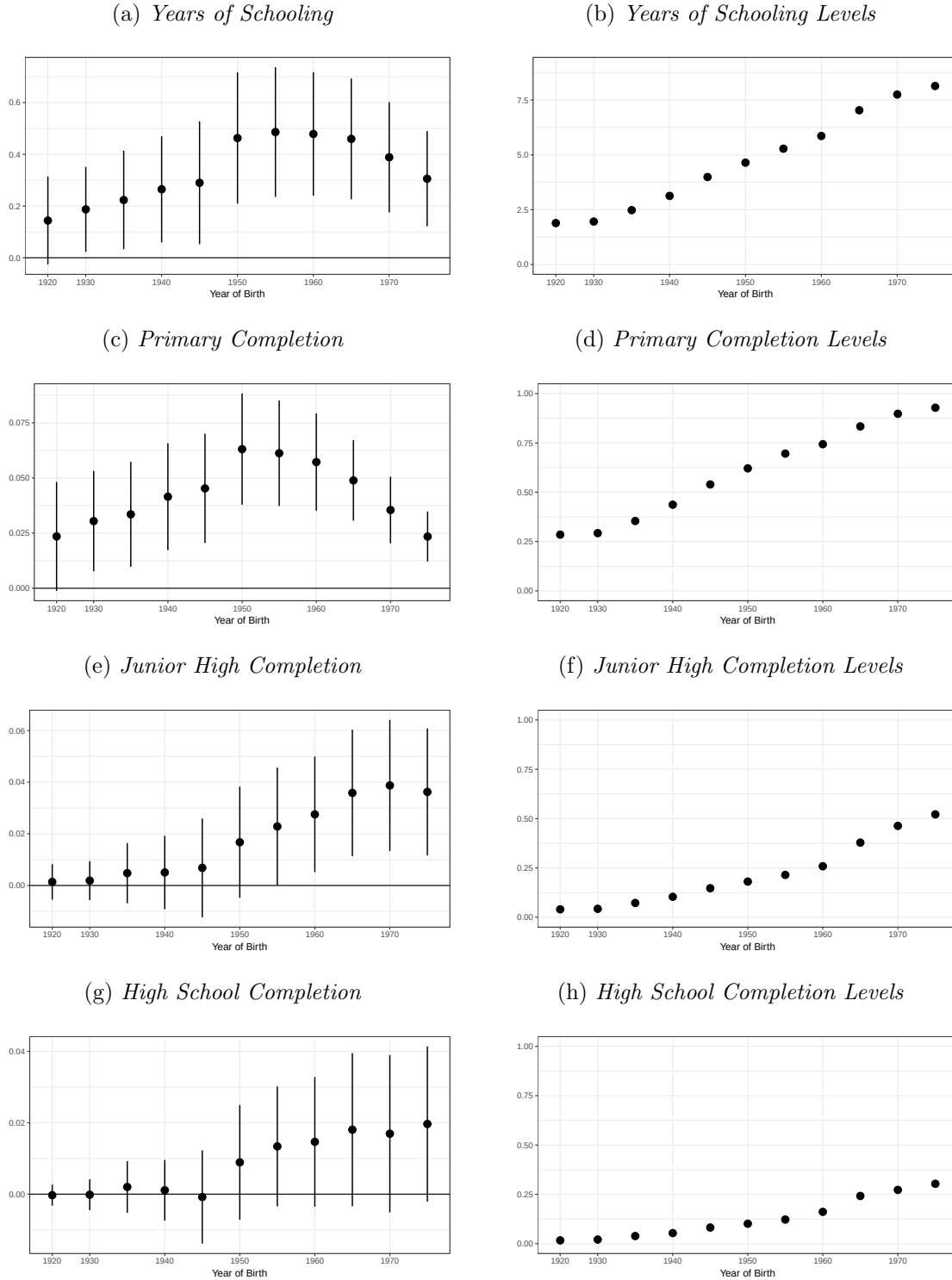
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure 10: Expenditure (2001-11)



Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure 11: Education by Cohort: Subjected Villages (2000 Census)



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table 1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.209 (0.774)***	-0.019 (0.014)	0.014 (0.164)	-0.012 (0.016)	0.132 (0.114)	0.001 (0.012)	0.074 (0.105)	0.059 (0.141)
Obs	4,209	4,209	4,205	4,209	4,209	4,209	4,209	4,209
Clusters	380	380	380	380	380	380	380	380
Mean	31.60	0.26	2.54	0.06	25.50	0.29	32.94	24.89

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table 2: Land Tenure: Subjected Villages

	Village Land 2003 Total Land	Land Share	Village Land 1980 Total Land	Land Share	99th Pctile ÷ 90th Pctile	90th Pctile ÷ 10th Pctile	90th Pctile ÷ 50th Pctile	50th Pctile ÷ 10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.304 (0.851)***	0.014 (0.004)***	3.412 (1.541)**	0.012 (0.005)**	0.016 (0.262)	13.543 (8.399)	0.361 (0.272)	1.851 (1.985)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

Notes: The unit of observation is the village. Regressions include geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table 3: Schools (1980): Subjected Villages

	Public Non-INPRES Buildings	Primary Teachers	INPRES Buildings	Primary Teachers	Junior High Schools	High Schools
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.010 (0.019)	0.171 (0.105)	-0.035 (0.020)*	-0.211 (0.081)***	0.020 (0.009)**	0.007 (0.006)
Obs	4,205	4,205	4,205	4,205	4,205	4,205
Clusters	380	380	380	380	380	380
Mean	0.43	2.81	0.36	1.37	0.06	0.02

Notes: The unit of observation is the village. Regressions include geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table 4: Education: Subjected Villages

	Years Education (1)	2000 Population Census Primary School (2)	Junior High (3)	High School (4)	1980 Census No School (5)	Primary School (6)	Village Head Years Education (7)	High School (8)
Cultivation	0.223 (0.075)***	0.025 (0.006)***	0.016 (0.007)**	0.007 (0.006)	-0.017 (0.008)**	-0.004 (0.010)	0.100 (0.086)	0.029 (0.013)**
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table 5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag. (1)	Manuf. (2)	Retail (3)	Ag. (4)	Manuf. (5)	Commerce (6)
Cultivation	-0.040 (0.013)***	0.029 (0.010)***	0.012 (0.007)*	-0.034 (0.021)	0.007 (0.012)	0.027 (0.010)***
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table 6: Firms, Population, and Consumption: Subjected Villages

	Num. Manuf. Firms (1)	Manuf. Emp. Share (2)	Log Population Density (2003) (3)	Log Population Density (1980) (4)	Log Equiv. Consumption (5)
Cultivation	21.600 (8.655)**	0.024 (0.018)	0.063 (0.035)*	0.062 (0.032)*	0.005 (0.011)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.