

The Development Effects Of The Extractive Colonial Economy: The Dutch Cultivation System In Java*

Melissa Dell and Benjamin A. Olken

Harvard University and MIT

October 2018

Abstract

Colonial powers typically organized economic activity in the colonies to maximize their economic returns. While the literature has emphasized long-run negative economic impacts via institutional quality, the changes in economic organization implemented to spur production historically could also directly influence economic organization in the long-run, exerting countervailing 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 economic activity in the long run, and we discuss the contexts where such effects are most likely to be important.

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. We thank Dave Donaldson, Rick Hornbeck, Nathan Nunn, Pablo Querubin, and seminar participants at Boston College, CIFAR, Columbia, NBER Summer Institute, Santa Clara, Stanford, UC Berkeley, and UC San Diego for helpful comments and suggestions. Finally, we thank Robert Parthesius for providing the data on VOC destinations. Contact email: melisadell@fas.harvard.edu, address: Harvard University Department of Economics, Cambridge MA 02138.

1 Introduction

Extractive colonial institutions are often thought to be an important reason why some places remain persistently poorer than others. 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). The persistence of such institutions can lower economic performance today.

Counterbalancing this, however, is the fact that colonial powers often established complex economic systems to create the surplus that they wanted to extract, potentially facilitating economic activity in the long run. In the case of agricultural extraction, crops needed to be grown and processed before being transported to the home country. This typically involved a reorganization of the indigenous economy, technology transfer, and the construction of processing and transport infrastructure. Examples include the large-scale sugar processing and transport infrastructure that the Japanese created in colonial Taiwan and the tea plantations, processing plants, and railroads that the British established in colonial India.

This paper examines the persistent local development effects of this type of extractive system - which combined coercive agricultural practices with large-scale domestic processing - by studying the Dutch Cultivation System (*Cultuurstelsel*) in Java. The Cultivation System is a particularly prominent example of colonial extraction. 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 colonial state forced peasants along Java's northern coast to cultivate sugar, which was then processed in nearby Dutch factories for subsequent export to Europe. The revenues extracted from this system made Java among the world's most lucrative colonies, at their peak accounting for over one-third of Dutch government revenue (Luiten van Zanden, 2010). Trade with the East Indies, supported by the Cultivation System, was an important contributor to overall Dutch economic growth in the first half of the 19th century (Mokyr, 1976, p.103, 108, 261). Sugar production continued after the abolition of the System but collapsed during the Great Depression, as Indonesia lacked protected markets, and today Indonesia is one of the world's largest sugar importers.

Prior to the Cultivation System, the Javanese economy was heavily specialized in rice cultivation, largely for local consumption. The Cultivation System required a substantial reorganization of economic life in order to operate. At its heart were 94 water-powered Dutch sugar factories, built in the Javanese countryside, which processed raw cane 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, it 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. Village officials within the catchment areas were politically empowered by the colonial government to make this happen.

This study examines the effects of the two main changes the Dutch made to the economy: 1) the creation of manufacturing in the previously agricultural heartland, through the construction of the sugar factories, and 2) the coerced reorganization of villages surrounding the factories to grow sugar and supply labor for the factories. It is worth underscoring that this study does not consider the impacts of colonialism more generally, an inherently speculative question since there is not a good counterfactual for what would have happened if the Dutch, or other colonial powers, had not shown up at all. Rather, our aim here is to elucidate whether places subjected to more intensive colonial extraction along these two important and common dimensions have experienced a differential development trajectory. We use two different empirical strategies for each of these two aims.

To examine the impact of the creation of the sugar factories, we combine a randomization inference type approach with economic insights that allow us to construct counterfactual spatial configurations of sugar processing. We exploit the fact that factory catchment areas were often adjacent, and the factories could not be too close since each required an adequately sized catchment area for cultivating sugarcane. While many locations were similarly suitable, once one factory was placed, it constrained the locations of other nearby factories, creating many possible equilibria for site selection (Salop, 1979). For example, if a factory was shifted by two kilometers upstream, the neighboring factories (and their neighbors and so forth) would also need to be shifted. Most of the villages producing sugarcane would remain the same, but the spatial configuration of factories within cane producing regions would differ.

We compare the impact of proximity to an actual factory to the effect of proximity to counterfactual factories under alternative feasible arrangements, averaged across 1,000 alternative configurations; p values are computed by comparing the actual effect to the distribution of counterfactual effects. A feasible counterfactual factory configuration consists of a suitable site for each sugar factory, with sites spaced far enough apart to have adequately sized catchment areas. Sites within alternative configurations are selected randomly from the set of suitable locations. We identify suitable locations by imposing the following requirements: 1) since state-of-the-art sugar 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 from the actual factory, and 2) the amount of nearby sugar suitable land must be similar to the amount of suitable land near actual factories. Moreover, counterfactual factories in each configuration must be spaced sufficiently far apart - relative

to the distribution of actual factories - and must be balanced geographically - i.e. the average counterfactual location within a configuration must be near the average actual factory location. Figure 1 illustrates this approach, and Figure 2 presents a heatmap showing where the counterfactual factories are concentrated. To separate the effects of Dutch extraction from modern sugar processing, we focus on colonial factories with no modern sugar factory nearby, though most results are similar if we consider the entire sample.

Geographic characteristics such as topography and hydrology are balanced across locations of varying proximity to historical factories, as compared to locations of varying proximity to placebo factories. Moreover, important pre-period characteristics are balanced: distance to pre-period residency capitals - the government and economic centers of historical Java - distance to ports used by the Dutch East India Company - the historical trading centers - and distance to the Great Post Road - the only pre-period major road in Java.

We document that the colonial sugar industry substantially transformed economic activity in contemporary Java. SUSENAS household surveys, collected between 2001 and 2011, show 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. The comparison to the distribution of estimated counterfactual effects implies a p-value of 0.001. Moreover, households living within a few kilometers of a historical factory location have per-capita consumption that is about 10 percent higher than those living more than 10 kilometers away (after subtracting out the small counterfactual effect).

The baseline coefficient estimates difference out the average impacts of proximity to counterfactual factories, which capture other ways that outcomes could vary with distance to locations suitable for sugar processing. The impacts of proximity to counterfactual locations are typically small, and we document that results are qualitatively similar if we dispense with the counterfactual analysis and simply run an OLS regression on distance to the nearest factory. The effects are moreover concentrated within a few kilometers and after that remain flat at zero when moving further from the factories. This suggests that the impacts do not result simply from a reallocation of economic activity from nearby areas, a scenario that would tend to lead to negative effects for these locations relative to even further places.

Input-output linkages, infrastructure investments, and human capital accumulation appear to be important channels of persistence. During the 1950s and 60s, development economists hypothesized that linkages between sectors provided a key mechanism for the propagation of structural change (Rasmussen, 1956; Myrdal, 1957; Hirschman, 1960), and the composition of industries in these areas indeed suggests agglomeration through input-

output linkages as an important mechanism.¹ While the Dutch claimed the high quality processed sugar, the factories were permitted to sell the low quality sugar, which was costly to transport, on local markets. 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 serve as direct or indirect inputs to sugar processing - include farm and capital machinery. Downstream industries - those that use processed sugar as an input - include most other food processing industries. Areas near historical factories have a higher share of employment in downstream industries, even when we restrict to locations with no modern sugar factory nearby. Although the original source of sugar disappeared, downstream industrial centers persisted, plausibly because of agglomeration of downstream producers and the endogenous concentration of final goods consumers near initial production sites.

A second important channel of persistence is transport infrastructure. In 1830, Javanese road infrastructure was nearly non-existent and consisted primarily of a single road built across the island for military purposes. By the early twentieth century, Java had the densest road and rail network in Asia (Booth, 2007, p. 80). This infrastructure was predominantly in regions involved in the Cultivation System, built to transport processed sugar to the ports. We show using our methodology that, indeed, villages located within a few kilometers of a historical factory had higher road and railroad density in 1900. In the absence of the System, this infrastructure would almost certainly not have been built elsewhere in Java, as the Dutch made these costly investments only because the extraction of a large surplus ensured a profitable return. This advantage in infrastructure has persisted: these villages were more likely to have a paved road in 1980, and today have a much higher density of intercity and local roads, as well as railroads, than places just a few kilometers further away.

Over time, more industrialized and connected places near historical factories may have been able to gain access to public goods more generally - i.e. because they could better afford them or because the returns to public goods used in industrial production were higher. This would plausibly reinforce the differences in income and industrial structure. Indeed, we find that villages near a historical factory were more likely to have electricity and a high school in 1980, a time when high schools were very rare. Using the complete 100% sample microdata from the 2000 Census, we document that people in villages within a few kilometers of colonial factories are more 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.

¹See Acemoglu et al. (2016); Baqaee (2015); Acemoglu et al. (2012); Carvalho (2009) for the role of input-output linkages in transmitting macroeconomic shocks and Liu (2017); Lane (2017); Bartelme and Gorodnichenko (2015) for their role in development. Ellison et al. (2010) show that they are important in generating contemporaneous industry co-agglomeration in the U.S.

The discussion thus far has focused on the establishment of the factories. The Cultivation System also subjected villages within a contiguous catchment area surrounding each factory to forced sugarcane cultivation, fundamentally altering their economic and political organization. To estimate these impacts, we use a different empirical design: a spatial discontinuity across the borders of the catchment areas. We obtained a 19th century handwritten list of the over 10,000 villages subjected to forced cultivation from Dutch archives in the Hague and computed the exact location of the Cultivation System boundaries by matching these villages with modern georeferenced locations (Figure 3). Inside, villages cultivated sugarcane for the government, whereas outside they did not. The boundaries form a multi-dimensional discontinuity in longitude-latitude space and allow us to control flexibly 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.

The Dutch did not have enough officials to directly manage the Cultivation System in the over 10,000 villages subjected and instead empowered Javanese village heads to manage it by giving them considerably greater command over land and labor than they had exercised previously. Village heads also received incentive payments proportional to their village's cane cultivation. The historical literature emphasizes that in response, village heads redistributed land to the village, where they would firmly control it. We therefore begin by examining whether the Cultivation System persistently influenced land allocations. Javanese villages typically have some amount of land set aside permanently for use by the village for public purposes. We find that there is about 10 to 15 percent more village public-use land in Cultivation System villages, in both 1980 and 2003.

To the extent that not all the revenues from these public lands were extracted for private gain, they may have helped villages to overcome the collective action barriers inherent in funding public goods. Indeed, anthropological evidence from the 1970s suggests that the practice of using village lands to provide public goods for the village was widespread in Central Java (Antlöv and Cederroth, 1994). We document that households in Cultivation System villages are more educated, and the impacts go all the way back to the 1920s cohort, which was educated during the Dutch period. Historically, villages that wanted a school needed to fund the school building themselves, and indeed we find that cultivation villages appear to have had more school infrastructure prior to the nationally-funded school construction program begun in the 1970s. These estimates isolate the impacts of being subjected to forced cultivation, and are above and beyond the effects of being closer to a factory, which the RD controls for flexibly. Differences in human capital plausibly impacted economic structure more broadly, as cultivation villages have a higher percentage of households working in manufacturing and retail and fewer in agriculture, both in 1980 and today. People living in

these villages are no poorer than those living just outside.

In sum, we have shown that the Cultivation System transformed the local economy in ways that have led to sustained higher levels of development today. In interpreting these effects, it is worth keeping in mind how limited economic activity was in rural Java in 1830, prior to the start of the System. The economy was primarily agricultural, landless peasants were customarily attached to land-owning households and hence not available for factory labor, and transport infrastructure and external finance were nearly non-existent. Moreover, village heads tended to be weak relative to landowners, and village level public goods provision was limited.² The Cultivation System altered these impediments to industrialization to create an extractable surplus. Importantly, the results do not imply that the Cultivation System was beneficial for the contemporaneous native population, and indeed it likely had large negative impacts on those who were coerced. However, the System also made major alterations to economic life to create an extractable surplus, and following independence the Javanese were able to utilize and expand the economic organization that had developed over the previous century to generate economic activity.

The conventional wisdom is that extractive institutions are a major determinant of poor long-run growth performance, both across countries (Acemoglu et al., 2002, 2001) and within countries (Dell, 2010; Lowes and Montero, 2016). This paper provides a prominent counter-example, from one of the largest and most lucrative colonies in the world, demonstrating that those areas where colonial extraction was more intense appear to be better off economically today. This raises the question of why the Dutch sugar system led to positive long-run development outcomes, in contrast to other well-studied examples, such as the *mita* in Peru (Dell, 2010) or rubber in the DRC (Lowes and Montero, 2016). While this is difficult to answer definitively, there are important aspects of the Dutch colonial experience that can help explain these differences and provide some guidance for interpreting other colonial episodes. Two factors in particular – the role of manufacturing and substantial investments in infrastructure – appear central.

First, the Cultivation System required substantial local manufacturing in order to process sugar cane prior to transport, and we document that these areas have persisted as manufacturing centers, despite the mass exodus of Dutch human and physical capital following independence and other interveing factors during the past two centuries. This was not true of the *mita*, where instead conscripts within the subjected area were marched from their communities to Potosí to mine raw silver that was then exported, nor of rubber in the Congo, which simply needed to be cut into cubes and dried before being exported to Europe. Linkages between sugar and other sectors plausibly provided a key mechanism for

²This type of economic and political organization was common across Southeast Asia. See for example Lieberman, 1993; Coedès, 1966; Mabbett, 1977; Kulke, 1986; Tambiah, 1977; Wolters, 1999; Tambiah, 2013.

the propagation of industrialization (Rasmussen, 1956; Myrdal, 1957; Hirschman, 1960).

Second, we show that the Dutch made important infrastructure investments in both rail and roads to transport the manufactured sugar, and these investments have persisted to the present. The literature estimates that colonial rail investments had important contemporaneous effects (see Donaldson (2010) on colonial India), and the Javanese context suggests that infrastructure effects are highly persistent. The investments in Java contrast to the Congo, where the Belgians had strong incentives to maximize short-run extraction and made few infrastructure investments, and to Peru, where silver was extracted from a single remote mine.³ While there is an important literature on colonial investments in areas such as health, education and infrastructure - with results showing substantial heterogeneity - this study differs in examining investments in extractive capacity, as opposed to investments that were potentially intended to benefit the local population.⁴

The study also contributes to the literature on persistence in urban geography. This literature documents a remarkable stability in the location of economic activity even following large-scale shocks such as bombing or international boundary changes (Davis and Weinstein, 2002, 2008; Brakman et al., 2004; Miguel and Roland, 2011; Redding et al., 2011). This is consistent with an important role for natural advantage or fixed capital, but substantial persistence appears even in contexts where these explanations are unlikely to be important. In particular, Bleakley and Lin (2012) document persistence of economic activity around historical U.S. portage sites, despite the striking fact that the practice of transporting goods via portage around falls in rivers disappeared over a century ago. They highlight agglomeration as a plausible explanation. In our context, the natural advantage of being on a river that can power sugar crushers has likewise long since disappeared (and in any case is differenced out by the counterfactuals), yet the effects of being near a historical factory location still persist. This study improves our understanding of spatial persistence by highlighting input-output linkages and transport networks as specific channels of persistence.

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

³Around the time of Belgian colonization, the invention of the inflatable tire greatly increased global rubber demand. The Congo was a rich source of indigenous rubber, but with time trees could be transplanted elsewhere, driving down prices and reducing extractive potential.

⁴Huillery (2009) finds that colonial investments in health, education, and infrastructure in West Africa positively influence long-run development, whereas Lowes and Montero (2017) document that historical health campaigns in French Central Africa lower trust in modern medicine. Cagé and Rueda (2016) show that Christian missions increase HIV prevalence in general but lower it where they made health investments. Gallego and Woodberry (2010) find that historical missionary activity in Africa increases education today, but only in areas where Protestant and Catholic missions competed, whereas Caicedo (2014) finds that colonial Jesuit missions increase human capital in Paraguay and Cagé and Rueda (2016) document positive impacts of the printing press on education and political participation in regions with missions. Jedwab and Moradi (2016) and Jedwab et al. (2015) provide evidence that colonial railroads persistently affected the distribution and aggregate level of economic activity in Africa.

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 An Overview of the Cultivation System

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 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).⁵ We focus on sugar cultivation, which took place on the populated plains of Java’s Northeast Coast. Coffee, while profitable, was primarily grown on forested mountain slopes that were uninhabited initially (Elson, 1994, p. 65). Extant data on coffee 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 use of coercion in sugar production was typical at this time, see Dippel et al. (2015) for a theoretical and empirical examination of how returns from sugar were invested in coercive institutions in the Caribbean. 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). The factories were primarily established in the 1830s and after that changed little. (Elson, 1984, p. 71). 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 workers labored in the sugar factories or related services (i.e. transporting cane and firewood to the factories) (Elson, 1994, p. 215).

Communities were forced to both produce raw sugar and provide conscripts for factory labor. Sugar cultivators never received a wage, but the number of free laborers in the factories expanded significantly across time. 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

⁵Indigo 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.

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 System.⁶ Outside Cultivation System villages - and on land not used for sugar in those villages - the economy was heavily specialized in rice cultivation.

The Cultivation System was primarily administered by Javanese officials, and village heads in the over 10,000 subjected villages were incentivized to cooperate with a variety of methods. While there was a threat of Dutch military coercion in the background, the use of military force on the equilibrium path was rare. Importantly, the Dutch provided financial incentives to cooperate, distributing payments (*kultuurporcenten*) proportional to the cane harvest through the village administrative structure. While there were formal rules about how these payments should be distributed to cultivators, in practice the village head often had leverage to do as he pleased, assuming cane quotas were met. Moreover, village heads were tasked with allocating land and labor for the System, with the potential to utilize this power for their own ends, such as ordering labor to cultivate their personal estate; (Van Niel, 1992, p. 139). Finally, the Dutch changed the system of village head selection to gain more control. Traditionally, the head's term was limited and required consent from landholders. With the advent of the System, 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, but the Dutch still maintained the right to replace non-cooperative heads. The combination of these forces enabled them to maintain control.

The Dutch phased out the Cultivation System in the 1880s, and sugar expanded rapidly 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.⁷ 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.

2.2 The Cultivation System's Historical Impacts

Relative to the economics literature on extractive institutions - which tends to emphasize persistent depressive effects on economic activity - the historical literature on the Cultivation System underscores its role in promoting modern exchange. Extensive road and rail infras-

⁶The exceptions were the indirectly ruled Principalities, which had private estate agriculture and were exempt from the Cultivation System.

⁷For example, the Philippines and Hawaii had protected access to the U.S., and Taiwan had protected access to Japan.

tructure were constructed to connect sugar producing regions to ports, plausibly promoting market integration (Elson, 1994, p. 251-252, Ricklefs, 2008, p. 158). Moreover, the 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. Third, 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.

Subjected communities also received incentive payments based on the amount of sugar produced, and historians argue that crop payments, by injecting currency into what had previously been a non-monetized subsistence economy, promoted modern exchange (Elson, 1994, p. 261, Ricklefs, 2008, p. 158). The payments often exceeded the land tax owed to the Dutch (Elson, 1994, p. 311, Van Niel, 1992), and residencies (provinces) more involved in the Cultivation System experienced a greater growth in markets while it was in force.⁸

Much of the historical literature on subjected villages has focused on the effects on land distribution, with a general consensus that the System increased landholdings of the village government. 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). A subsequent literature likewise hypothesizes that the Cultivation System had equalizing effects, through increasing the amount of communal village land or through leading village land to be divided more equally.⁹ A revisionist literature, initiated by Elson (1994, 1984), similarly emphasizes an increase in village collective landholdings, but argues instead that this empowered village elites (see also Ricklefs (2008, p. 159)). In particular, village heads were responsible for assigning land for government cultivation, as the Dutch relied on Javanese authorities to ensure that resources were devoted to government cultivation, and the Dutch distributed incentive payments proportional to sugar cultivation to the village head. Village heads plausibly responded by redistributing land to the village (where they controlled it) or to their cronies (Van Niel, 1992, p. 139). By 1857, 20% of all the irrigated land in Probolinggo Residency - a center of sugar cultivation - was controlled 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).

⁸Government sugar land was assessed at the highest rate, so if anything increased the village's tax burden.

⁹See Elson (1994, p. 162) for a review.

2.3 Other Historical Examples

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.

3 Data

3.1 The Cultivation System

Data on the Cultivation System are drawn from handwritten 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 in 1853 the Minister of Colonies appointed the Umbgrove Commission to conduct an inquiry into government sugar cultivation. They spent the next four years collecting detailed data. Handwritten documents list which villages contributed to each sugar factory and how much land and labor each village provided, for the approximately 10,000 subjected villages. Appendix Table A-1 provides an example. The manuscripts also contain qualitative information about the social situation and sometimes show sketches of the catchment areas.

Historical villages are matched with coordinates from the U.S. National Geospatial Intelligence Agency's Geonames database, which provides a detailed list of populated places, including many sub-village units.¹⁰ Matches are made using the historical village's name and the location of its historical district. The manuscripts also list the distance between the village's fields and the factory, which allows us to distinguish between multiple matches. We match 6,383 historical villages with coordinates, which are located in 2,519 modern village polygons. Modern villages are small - there are over 30,000 in Java. The factories are matched using the same procedure. The combination of the historical sub-village, plus the fact that the factory had to be adjacent to a river, means that they can be precisely located.

¹⁰Data are available for download at <http://geonames.nga.mil/gns/html/index.html>.

Not all of the approximately 10,000 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 4, which shows the factories (stars), matched village points (dots), straight lines connecting the factories to matched points, and catchment boundaries (thick lines) for an example catchment area. Modern village polygons are shown using thin lines 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.¹¹

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 (Table A-1). All contain village identifiers that we match to modern village boundaries.¹²

The variation that we exploit is extremely fine, and anything above the village level aggregates it away. Villages are very small - with over 30,000 in Java today - and while a number of older data sources exist, they are too aggregated to be useful for our analysis. We can though trace historical effects using cohort analysis and detailed infrastructure maps published in 1900 by the Dutch Topographic Bureau (*Topographisch Bureau*) in Batavia.

4 Impacts of Sugar Factories

4.1 Empirical Strategy

4.1.1 Assignment to Sugar Cultivation

Understanding why sugar factories were located in particular places is central to identifying the long-run impacts of the Cultivation System. At the start of the System, there was little

¹¹The 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.

¹²We use the term village to 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.

sugar production in Java.¹³ The government sugar factories we study in this paper were created mostly in the 1830s and remained essentially fixed thereafter.

The Dutch had limited information about rural Java in 1830, as no agronomic or population surveys existed (Ricklefs, 2008). The past decades had been devoted to costly wars of subjugation and not administration. During the 1830s, the colonial government required Dutch officials - who had a limited presence in the countryside - to submit reports on the suitability of their areas for various export crops, and these were used to locate government cultivation. Van Schaik (1986, p. 47-49) analyzes extant reports held by local archives for two of the main sugar producing residencies - Tegal and Pasuruan - which suggest that Dutch decisions were based on fairly coarse knowledge.

Specifically, several factors were cited as important in determining where to site factories: 1) sufficient surrounding land suitable for growing sugar, 2) proximity to a river (to generate power for the factory), 3) proximity to firewood, transport routes, and population, and 4) sufficient spacing between factories to ensure that each factory had enough land nearby that could grow cane. We discuss these in turn.

First, 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. Each factory was allocated a certain area from which to draw the land and labor it required, and these subjected villages were typically within four to seven kilometers of their factory. This led almost all factories to be located on the plains of northern Java. However, the Dutch lacked the agronomic surveys and experience cultivating cane that would have been necessary to further fine-tune cultivation locations. Indeed, long-run average yields under government cultivation, summarized in the Dutch annual reports on the System held in the Hague, subsequently revealed that subjected areas had a wide range of sugar yields.

Second, proximity to a river was crucial for locating factories because sugar processing used iron crushing cylinders powered by water wheels.¹⁴ Since sugar had to be grown in flat, low elevation plains, this kept terrain variation along suitable rivers to a minimum.

Proximity to transport, population, and timber were also considered. We will show that within our sample proximity to pre-existing ports, road infrastructure, and cities is balanced. This suggests that constraints on remoteness operated at a more aggregate level to avoid much more remote regions that are not in our sample, rather than to fine-tune locations

¹³During 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 the 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).

¹⁴It was not until the late 19th century that steam power became dominant, and indeed after the abolition of the System there was a large-scale expansion of cultivation to areas where processing would not have been possible prior to steam power.

within our sample area.

Factories could not be located in areas where transport costs were prohibitive. This was the other reason that most factories were located in northern Java, where the only pre-existing trans-Java road was located, but distance to this road is balanced in our sample.¹⁵

Data do not exist to check whether pre-period population varies with proximity to historical factories for the over 10,000 historical villages nearby, as the Dutch did not conduct a census of village populations until 1930. However, citing Dutch correspondence, Van Schaik (1986, p. 49) writes that only in the late 1840s – after the establishment of the factories we study here – was it realized that population should have been “taken more seriously into consideration” in locating government sugar cultivation. Moreover, our empirical strategy will compare actual factories to nearby counterfactual locations that likewise tended to be surrounded by villages that contributed sugar cane to the System. Even if the Dutch tended to subject areas with more population to forced cultivation, the counterfactual factories also are typically surrounded by these areas.

Finally, with regards to forests, maps systematically documenting forest cover do not exist prior to the late 19th century, and forest cover likely changed substantially over time due to forest exploitation. Hence, we cannot reconstruct exactly what forest cover looked like as of the early 1830s. However, Dutch reports on the siting of sugar factories indicate that forest cover at the time was dense enough that it was not generally a constraint (Van Schaik, 1986, p. 47).

Within the sugar suitable plains, sufficient spacing was a dominant constraint on factory placement. Catchment areas tended to be adjacent to each other and avoided major pre-existing cities.¹⁶ Critically, once one factory was sited, another one could not be located too close even if there were many suitable places along a river, since the catchment areas were adjacent, and each factory required a sufficiently sized catchment area to meet the minimum scale of production. We will exploit this fact to develop an empirical strategy for identifying the long-run effects of constructing a sugar processing infrastructure.

4.1.2 Empirical Specification

Estimating the effects of proximity to a historical factory requires identifying a plausible set of counterfactual factory locations. We exploit the fact that there were many possible equilibrium factory configurations, since catchment areas were typically adjacent and each needed to produce enough cane for the factory to operate. A variety of locations along rivers

¹⁵The primary non-remote region that is sugar suitable but did not cultivate sugar is the Dutch residency of Rembang, where the Dutch were cultivating tobacco instead for the Cultivation System.

¹⁶Colonial Java was divided into 24 residencies, which were the center of colonial administration and the major cities of the time. The colonial residency capitals by and large remain the largest cities today.

between actual factories would have been suitable - and would not have resulted in major overall changes in the set of villages that produced raw cane - but factories were not placed in these locations due to the spacing constraints.

A feasible spatial equilibrium for the Cultivation System’s factories consists of a suitable site for each factory, with sites spaced far enough apart to have adequately sized catchment areas. In the baseline specification, we identify feasible counterfactual factory sites by imposing the following requirements, and an extensive set of appendices documents robustness:

1. Since the state-of-the-art sugar processing technology was water-powered, sugar factories were located along rivers; we therefore only consider sites that can be reached by moving 5 to 20 kilometers upstream or downstream via river from the actual factory.
2. 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 the pre-determined component of 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 elevation and slope distributions to be highly skewed.¹⁷
3. Counterfactual factories for each draw need to be at least as far apart as the 10th percentile in the distribution of actual factories.
4. Each set of counterfactual factory draws must be geographically symmetric on average around the actual factories. In other words, the average latitude and longitude of the counterfactual factories must be similar to the average latitude and longitude of the actual factories.¹⁸

This approach is illustrated in Figure 1. 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

¹⁷An alternative would be to use FAO suitability data, but unfortunately these are too aggregated to be useful when exploiting village level variation.

¹⁸We achieve this balance using a simulated annealing procedure that minimizes the average distance between the counterfactual coordinates and the actual factory coordinates. There are by construction many combinations of counterfactual factories that minimize this distance, and our procedure selects 1000 of them to serve as counterfactuals. Within each counterfactual, we conduct J iterations to locate a minimum. If the average distance between the counterfactual factories and the actual factories $d_j < d_{j-1}$, the counterfactual locations from iteration j is accepted. Otherwise, it is rejected with probability p . Following standard simulated annealing procedure, p declines from 0.25 to 0 as the algorithm proceeds. In practice, we achieve qualitatively similar results omitting this annealing step - see Appendix J.

suitable land, as compared to the distribution of actual factories. Panel c) shows a larger set of counterfactual factories. Figure 2 presents a heatmap showing where the resulting counterfactual factories are concentrated.

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 (counterfactual) 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 includes elevation, slope, distance to the coast, distance to the nearest river, flow accumulation, distance to the nearest 1830 residency capital, distance to the (pre-period) Great Post Road, and distance to the nearest (pre-period) Dutch East India Company (VOC) port. The $fact_v^j$ are nearest factory fixed effects, which ensure that villages are compared to other villages near the same (counterfactual) factory.

The point estimate of the effect of being distance i away from the factory is given by the difference between the $dfact_v^i$ coefficient for the actual factory and the average of the 1,000 counterfactual $dfact_v^i$ coefficients. If there are unobserved factors that change with proximity to locations suitable for factories, subtracting the mean will purge their influence. However, a potential drawback is that if counterfactual factories are placed too close to the actual factories - relative to the rate at which the effects decline to zero - the effects of the actual factories could contaminate the counterfactual means. Appendix B shows that this bias will be most pronounced for the bins closest to the factories, tending to make the pattern of estimated effects flatter and to bias coefficients towards zero. Robustness and simulation evidence outlined in Appendix B suggest that this bias is not quantitatively important in our setting, as effects decline sufficiently quickly relative to the locations of the counterfactuals.¹⁹

Another potential drawback of this method is that it inherently involves making choices about what constitutes a plausible counterfactual, and one may worry about robustness to different choices. Section 4.3 documents that estimates are highly robust to these decisions.

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. Appendix C shows that this method is correctly sized.

¹⁹ Of course, it could still be the case that areas of Java far from the factories are affected by the Cultivation System - i.e. by market access type effects from the infrastructure constructed to reach the factories (i.e. as in Donaldson and Hornbeck (2016)) - a fascinating question beyond the scope of the current study.

4.1.3 Pre-Characteristic Balance

We begin by examining balance in geographic and pre-period characteristics. These characteristics could vary with proximity to actual factories, but we would expect this variation to be similar when comparing actual and counterfactual factories, if indeed the counterfactual locations are plausible.

We illustrate the patterns in the data by plotting the difference between the $dfact^i$ coefficients for proximity to the actual factories and the mean $dfact^i$ coefficients for proximity to the counterfactual factories. As described above, we denote the significance of each of the $dfact^i$ coefficients relative to the counterfactual $dfact^i$ distributions. Crosses indicate coefficients that are above the 95th percentile of the counterfactual distributions, solid dots denote coefficients above the 90th percentile, and hollow dots indicate coefficients below the 90th percentile. The figures highlight the general shape of the relationship by plotting a linear spline with kink points every 3km.

Figure 5 documents geographic balance. The characteristics that we consider are elevation, slope, distance to the coast, flow accumulation, 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. We also examine important pre-period economic variables: distance to the nearest 1830 residency capital - the political and economic centers of historical Java - distance to the Great Post Road - the only major road in Java at the start of the Cultivation System - and distance to pre-period (1595-1660) Dutch East India Company (VOC) ports - the historical trading centers of Java.

There are not consistent relationships between these characteristics and distance to the nearest factory, suggesting that the identification assumptions are reasonable. Moreover, the magnitudes of the distance to factory coefficients are small – for example, moving from 1km away from a factory location to 10km away is associated with a statistically insignificant difference in elevation of about 2 meters.

4.2 Results

4.2.1 Economic Structure

We now turn to an examination of whether government sugar factories influence industrialization in the long-run, even in places where sugar has not been processed for many decades. We hence focus here on places without a modern sugar factory nearby, in order to isolate effects that do not go directly through the persistence of sugar processing. Subsequently, we will explicitly examine modern sugar production.

We begin by illustrating the methodology in more detail, using as a dependent variable an indicator for whether the individual works in the agricultural sector, taken from the SUSENAS 2001-2011 household surveys.²⁰ The sample includes prime age males aged 18 to 55 to avoid confounding labor market participation.²¹ For each factory we take 1,000 independent random draws that shift the factory to a suitable counterfactual location, constraining the factories in each counterfactual spatial arrangement to be sufficiently far apart. We estimate equation (1) using distance to the nearest actual factory, as well as distance to the nearest counterfactual factory in each of the 1,000 counterfactual configurations.

The sub-plots in Figure 6, panel a) show the counterfactual distributions of absolute coefficients for each of the $d\text{fact}^i$ 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, which measures 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 $d\text{fact}^i$ coefficients for proximity to the actual factories, subtracting the means of the counterfactual distributions shown in panel a). Crosses indicate coefficients above the 95th percentile of the counterfactual distributions, solid dots denote coefficients above the 90th percentile, and hollow dots indicate coefficients below the 90th percentile. These figures also plot a linear spline with kink points every 3 kilometers.

Individuals within a few kilometers of a historical factory are around 18 percentage points less likely to work in agriculture than those 10 to 20 kilometers away, relative to the effects on the counterfactual factories. Appendix Figure G-2 shows a version of this plot where the mean of each counterfactual $d\text{fact}^i$ distribution has not been subtracted from each coefficient, and the estimated effects are similar.

Throughout the analysis, we will see that results take this hockey stick shape: large in the immediate vicinity of the factory and then dissipating and remaining flat. This suggests that the impacts around the factories do not result simply from a reallocation of economic activity from nearby areas, a scenario that would tend to lead to negative effects for the latter locations relative to even further places or the counterfactuals.

²⁰When 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.

²¹Estimates, available upon request, are similar using all individuals.

We also conduct an exercise that shifts all the historical factories up or down the river by the same distance and then estimates equation (1) for each of these common shifts. Each sub-plot in panel c) shows the γ_i coefficients for a $d\text{fact}^i$ bin in distance to the nearest factory: 0-1 km, 1-2 km, etc. The x-axis plots the magnitude of the common shift upstream (left side of the plot) or downstream (right side of the plot), in kilometers. The y-axis plots the γ_i , with the length of each bar indicating the magnitude of the counterfactual estimate. 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 counterfactual factories are moved away from the actual factories. We can construct counterfactual means by subtracting the average of these counterfactual distributions, leaving out shifts within 5 kilometers upstream or downstream of the actual factories to minimize counterfactual means being contaminated by actual effects. Panel d) is analogous to panel b). The patterns that emerge using this set of counterfactuals are broadly similar, with the modestly lower coefficients plausibly explained by actual effects contaminating the counterfactual means.²²

Figure 7 examines the industrial structure near government sugar factories in more detail, examining both different sectors and different time periods. Due to space constraints, only the coefficient plots estimated using the counterfactual distributions from the independent random shifts are shown in the main text. The complete set of counterfactual distributions for the independent and common shifts are shown in Appendices D and E, respectively. Appendix H reports coefficient plots constructed using the common shifts.

For comparison, 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.²³ 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 25 to 30 percentage points less agricultural than those just ten to twenty kilometers away. Indonesia was more agricultural in 1980, and these results suggest some convergence with more recent industrialization.

Next, panels c) and d) consider employment in manufacturing, again using data from SUSENAS and the 1980 Population Census. The pattern for manufacturing during 2001-2011 is the inverse of that for agriculture, with around six percentage points more individuals working in manufacturing in the immediate vicinity of historical factories relative to further away. The 1980 Census likewise reveals that manufacturing employment is higher within

²²If the river branches upstream or downstream - which is common - we choose which branch to locate the counterfactual on at random. Hence, counterfactuals near the historical factories are over-represented in this sample relative to when we select counterfactuals using iid random draws. The bias described in Appendix B becomes larger, making these coefficients somewhat smaller. We nevertheless use the 5 kilometer upstream/downstream from the factory restriction - as in the independent shifts exercise - to be consistent.

²³We 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 conditions.

a few kilometers of a historical factory. 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 worked in manufacturing.

Third, we examine employment in retail. For 2001-2011, places in the immediate vicinity of a factory have around 9 percentage points more employment in retail relative to places ten to twenty kilometers away, and the coefficient for the closest bin is above the 95th percentile of the counterfactual distribution (panel e). In contrast to manufacturing, the retail effects are somewhat higher today than in 1980.

These occupational patterns – more manufacturing, more retail, and less agriculture – are consistent with the areas near historical factories being more urban. Panels g and h examine log population density in 2003 and 1980 (respectively), defined as the log of population in each village per square kilometer. Areas within 1 kilometer of a factory are much more densely populated than areas 10 kilometers away in both 1980 and today.

The effects relative to the 20 kilometer bin flatten out within a few kilometers, indicating that the population response is not drawn disproportionately from places somewhat closer to the historical factories relative to places further away.²⁴ Of course, if population is drawn from a large area, an agglomeration shadow, i.e. a decline in population reflecting reallocation to the area of agglomeration, could exist but be undetectable. Historical vital statistics data do not exist to disentangle an endogenous fertility response from a migration response, but as we will see the Cultivation System also had other impacts that cannot be solely explained by the reallocation of fixed factors.

4.2.2 Sugar and Linked Industries

The above results suggest that historical government sugar factories impact current economic activity through agglomeration externalities, and we now examine this mechanism in more detail. Firms that used sugar as an input may have been started near historical sugar processing plants - creating additional value through input-output, labor skill, and technological linkages (Ellison et al., 2010; Marshall, 1890). Final goods consumers or workers may have also clustered near historical production centers, with these patterns persisting even after the original sugar factories disappeared. In the absence of government sugar factories it is unlikely that similar agglomeration centers would have arisen elsewhere, as a lack of access to capital and the feudal-like organization of the rural economy more generally impeded the formation of modern sugar processing infrastructure.

²⁴ The typical medium-sized city in Java has a radius of less than five kilometers, so the local effects that we document are what we would expect if the Cultivation System converted areas around historical factories in rural Java into medium-sized cities.

Before examining linkages in detail, we consider whether sugar production itself has persisted. Figure 8, panel a) shows - using the full sample of historical factories and 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.²⁵ In contrast, panel b) documents that once we drop the 19% of historical factories within 2 kilometers of a modern sugar factory - as we do in all of our baseline results - the relationship between distance and modern sugar production flattens out.

Panel c) shows that there is little relationship in the full sample between raw cane production, drawn from the Podes 2003 village census, 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 the restricted sample (panel d).

Input-output linkages were plausibly an important driver of agglomeration around sugar factories historically, since the low grade sugar that factories sold on local markets was costly to transport. To test whether such agglomeration has persisted, we construct weighted average employment shares for narrow manufacturing industries upstream and downstream from sugar processing.²⁶ 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. Suppose sector k uses inputs from sector j , which uses inputs from sector i . The Leontief inverse measures how much of sector i 's output sector k uses, both directly and indirectly (via sector j). The Leontief weights are described in more detail in Appendix Q.

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 we focus on manufacturing.) The main manufacturing industries downstream are in food processing, as sugar is an additive to many other 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. These downstream linkages could contribute to the retail impacts documented above.

Figure 8, panel e) shows that when all historical factories are included in the analysis, the weighted average share of employment in upstream manufacturing industries is about

²⁵The Economic Census captures all large manufacturing firms and a random sample of small and informal firms.

²⁶Sectoral employment in manufacturing is drawn from the 2006 Economic Census and total employment is from the 2000 Population Census.

three times higher 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, even when we limit the sample to historical factories that are not near modern ones, employment in manufacturing industries downstream from sugar is much higher near the historical factories (panel h). This is particularly true for places within 0 to 1 kilometer of a historical factory, and the effect is large. The difference in the weighted average employment share of downstream industries between the 0-1 km bin and places 10 to 20 km away is approximately equal to the sample mean. 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.

4.2.3 Transport Infrastructure

The analysis thus far has focused on the private sector, but public investments may also be an important channel of persistence. The historical literature emphasizes that the Dutch government constructed road and rail networks to transport sugar to ports. The Dutch made large infrastructure investments precisely because it was profitable for them due to the extraction of a surplus, and they would have been very unlikely to make these investments elsewhere in the absence of extraction.

Figure 9, panels a) and b) examine colonial road and rail density in 1900. The data are from a detailed map that we geo-referenced of roads and railways published in 1900 by the Dutch Topographic Bureau (*Topographisch Bureau*) in Batavia. In the analysis of infrastructure - and for the remainder of the study - we drop historical factories near modern sugar factories to isolate impacts that do not go directly through sugar production, with Appendix K documenting similar effects when these places are included.

Both colonial road and rail density are substantially higher near historical sugar factories, and these effects are atypical relative to the counterfactual distributions. When comparing the 0-1 km bin to places 5-20 km away, the difference in colonial railroad density is more than twice as large as the sample mean. The difference in road density is also substantial.

Panels c) through f) document that these effects have persisted through the present, plausibly playing an important role in generating long-run economic impacts. Data from the 1980 Podes, a census of village governments, show that villages in the immediate vicinity of the historical factories were less likely to only be accessible via a dirt road historically (panel c). Today, both intercity and local road density are higher near historical factories. When

comparing the 0-1 km bin to places 5-20 km away, the difference in intercity road density is about twice as large as the sample mean (panel d). Differences in local road density (panel e) and railroad density (panel f) today are likewise large. Infrastructure and input-output agglomeration mechanisms could plausibly reinforce each other.

4.2.4 Other Public Goods

Over time the more industrialized and integrated areas near historical factories may have gained better access to public goods more generally, because they could better afford or lobby for them, or because the returns to public goods used in industrial production were higher. Figure 10 examines two of the main public goods provided in Java: electricity and schooling. 1980 Podes data reveal that places in the immediate vicinity of the historical factories were about 45 percentage points more likely to have electricity than places ten to twenty kilometers away (panel a). The effects fall far in the right tails of the counterfactual distributions. Moreover, we find that in 1980, areas in the immediate vicinity of a factory were four percentage points more likely to have a high school than places ten to twenty kilometers away, relative to a sample mean of only two percent of villages with high schools (panel b). While high schools are more common today, the effect remains (panel c). High schools were mostly built in administrative centers, and 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 d).

Figure 11 examines long-run impacts on schooling in more detail. 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 2000.²⁷ We focus in the main text on three representative cohorts: 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). Appendix R shows results for the complete set of cohorts.

Panel a) documents that individuals within the vicinity of a historical factory have around a year more schooling than those located ten to twenty kilometers away, relative to a sample mean of 4.9 years of schooling. 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

²⁷We use ten year age groupings to increase power for the oldest cohorts, since these are much smaller.

when primary access was far from universal. High school completion again shows a similar pattern (panels e and f). Effects are largest for 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. We focus here on the public provision of human capital through schooling, but it is also worth noting that historical factories may have increased specific dimensions of intangible human capital that are relevant to modern production - i.e. by exposing workers to modern factory methods.

4.2.5 Household Consumption

We conclude by examining effects on household consumption, using pooled data from the SUSENAS household survey. 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. Figure 12 shows that consumption levels in areas immediately adjacent to the historical factories are around 10 percent higher than areas 5 kilometers further away. Consistent with these being relatively integrated areas, this effect is about what we would expect given the differences in education. People living adjacent to a historical factory have about 1.25 more years of education, and an 8 percent return to schooling (Duflo, 2001) would yield the observed consumption differences.

4.3 Robustness

The counterfactual factory analysis requires a series of decisions about what defines a plausible counterfactual location, and Appendices F - P show that estimates are highly robust to these choices. It is useful to first highlight the underlying reasons for this robustness - which provide intuition for why the results are unlikely to disappear regardless of these decisions - as exploring every possible combination of robustness checks (i.e. running robustness checks on the robustness) would generate a very large number of coefficients.

Effects are pronounced at the exact locations of the historical sugar factories and decline steeply moving away from the factories. The outcomes otherwise vary more smoothly across space, even near the limited sets of locations that are plausible alternative factory locations. Thus - while we do not have a perfect model of the Dutch factory location decision - for an omitted variable to drive our results it would need to closely mimic the spatial pattern of the actual factories, without being present at the counterfactual sites. Moreover, an omitted variable would need to generate the specific types of effects we find, such as specialization in industries downstream from sugar processing. It is hard to think of an omitted variable that is a more plausible candidate than the presence of a historical sugar factory.

When we dispense with all the choices inherent in the counterfactual factory analysis and run an OLS regression on distance to the nearest factory with standard errors clustered by $100km^2$ grid cells, estimates are very similar (Appendix F). Relatedly, if we do not subtract the counterfactual means but do use the counterfactual approach to compute the standard errors, estimates are again similar to the baseline (Appendix G).

Next, Appendices H - L consider variations to the placement of counterfactual factories. These robustness checks illustrate that counterfactual means tend to be small - and actual effects atypical relative to the counterfactual distributions - regardless of how we choose the locations of the counterfactual factories. We first consider a completely different method for locating the counterfactuals. Instead of randomly allocating counterfactuals to suitable locations, we shift all the factories up or down the river by the same distance, as described in Section 4.2.1. Appendix H shows that estimates are broadly similar.

The fact that the counterfactual means tend to be near zero suggests that the counterfactual factories are not contaminated by the effects of the actual factories (see Appendix B for a detailed discussion), but another check on this is to require greater spacing between real and counterfactual factories. Appendix I requires the counterfactual factories to be at least 5 kilometers away from the nearest actual factory in straight line distance. This cuts the sample of factories with suitable placebos nearby in half, yet estimates remain similar.²⁸

In the baseline, we focus on the set of factories that have nearby sugar suitable land and where counterfactual factory locations can be geographically balanced around the factory - i.e. the latter is not possible if sugar suitable land is only located on one side of the factory.²⁹ Appendix J documents similar estimates when we do not require the counterfactual set of factories to be symmetric on average around the actual factories. Moreover, the baseline - except where explicitly noted - does not include historical factories that are near modern sugar factories, as the focus is on effects that do not go directly through the persistence of sugar production. Appendix K shows that most conclusions do not change when these factories are included. Appendix L shows robustness to removing factories in close proximity to residency capitals.

Next, Appendices M - N consider modifications to the sample of villages included in the analysis. In the baseline, the sample includes villages up to 20 kilometers from the nearest historical factory, and Appendix M documents that similar patterns obtain when only villages within 10 kilometers of the factories are included. The baseline specification limits the sample to sugar suitable villages, in order to ensure that the sample is geographically similar and because the analysis of the impacts of being a subjected sugar village will focus on this

²⁸Results, available upon request, are likewise robust to using other minimum distance thresholds.

²⁹To be consistent with the analysis of subjected villages, we also exclude places located in cities that were pre-period residency capitals.

set of places. The effects are very similar when we include all villages in the sample, as documented in Appendix N, suggesting that the channels of persistence have little to do with sugar suitability.

Finally, we consider modifications to the empirical specification used to estimate the impacts of distance to actual and counterfactual factories. Appendix O shows that it makes little difference if we omit the nearest factory fixed effects. Appendix P documents that results are similar when we control for whether each village contributed land to forced sugar cultivation.

4.4 Discussion

In summary, the Dutch colonial sugar industry substantially transformed economic activity in rural Java in the long run. 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, these industrial centers persisted, sourcing sugar from elsewhere. The post-colonial setting might if anything be a setting where we would expect these effects to be less pronounced. For example, one might have hypothesized that industrial benefits would have accrued entirely to a small number of colonial elites or that any effects would have dissipated following the massive exodus of Dutch human and financial capital from Indonesia following independence. Even in the face of such major disruptions, however, input-output linkages appear to generate a substantial degree of persistence.

Moreover, infrastructure had to be constructed to transport Cultivation System sugar to ports, and the infrastructure remained long after the Dutch left. Since maximizing colonial profits was central to the Dutch objective, it is hard to imagine a scenario where they would have made these investments in manufacturing and transport infrastructure in the absence of a large extractable surplus.

Of course, some industrial production centers would likely have ultimately arisen in rural Java in the absence of the Cultivation System. However, given that rural Java in 1830 was a heavily agricultural economy governed by traditional agricultural labor norms in which landless peasants were customarily attached to landowning households, most production was for subsistence, and financial markets were largely absent, there were many barriers to overcome. This study highlights how colonial extraction can have a range of effects, some of which may promote economic activity in the long run relative to the counterfactual.

5 Impacts on Subjected Villages

5.1 Empirical Strategy

The discussion thus far has focused on the establishment of sugar factories. We next turn 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 3). Inside, 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:

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

where out_v is an outcome in village v . $cultivation_v$ is an indicator equal to 1 if the village grew 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.³⁰ The individual catchments around each factory form 17 larger contiguous catchment areas, and we estimate the RD polynomial separately for each one.³¹

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 functional form captures the impacts of proximity to a factory well. \mathbf{X}_v contains the same exogenous geographic characteristics that we used in the distance to factory analysis: elevation, slope, distance to the coast, distance to the closest natural harbor, distance to the nearest river, log flow accumulation, distance to the nearest 1830 residency capital, distance to the nearest VOC port, and distance to the Great Post Road. 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 baseline specification limits the sample to villages within 10 kilometers of the threshold. Standard errors are clustered at the sub-district level. Villages where sugar could not grow are not a suitable counterfactual. Hence, we limit the sample to areas that are suitable for sugar, where suitability is defined using the same elevation and

³⁰Regressions use a triangular kernel such that the weight given to each observation decays with distance from the threshold.

³¹Due to power limitations, we do not estimate the RD polynomial separately on either side of the threshold, as it is already extremely flexible. We show in the appendix for a lower dimensional polynomial that results are indeed robust to estimating it separately on either side of the threshold.

slope requirements described in Section 4. For the same reason, we exclude places in cities that were residency capitals.

RD plots are shown in Appendix S. Legibly showing multi-dimensional RD graphs where latitude and longitude are the running variables, as in Dell (2010), would require zooming in separately for each of the catchment areas, generating a very large number of plots. In the interest of concision, we instead show single dimensional RD graphs where the running variable is distance to the nearest catchment boundary. Negative values are used for places outside the catchment areas.³²

5.1.1 Pre-characteristic Balance

The key regression discontinuity identifying assumption is that all relevant factors besides treatment vary smoothly at the Cultivation System boundaries. This assumption is needed for observations located just outside the catchment areas to be an appropriate counterfactual for observations located just inside.

To assess the plausibility of this assumption, Table 1 examines a variety of geographic characteristics, using regressions of the form described in equation (2). The unit of observation is the village.³³ Column (1) examines elevation. The point estimate on *cultivation* is negative and statistically significant but is only 2 meters, a small difference. Indeed, the entire sample is in the plains, close to sea level (the mean elevation in the sample is 31 meters). In some of the alternative specifications examined in the appendices, this difference is smaller and not statistically significant, yet other results remain similar. Slope; flow accumulation; a dummy for being on the coast; and distance to the coast, nearest river, nearest 1830 residency capital, and the Great Post Road are balanced. There is a statistically significant difference in distance to the nearest VOC port - subjected villages are if anything further - though given the number of characteristics examined this difference may arise by chance.³⁴

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

³²Border segment fixed effects, geographic and pre-period controls, a spline in distance to the nearest sugar factory, and a linear polynomial in latitude and longitude - estimated separately by catchment area - have been partialled out.

³³Appendix Table AF-1 shows that results change little when it is a grid cell centroid.

³⁴We do not examine characteristics such as weather or soil quality since the lower resolution of these data mean treated and non-treated villages typically fall in the same cell.

established village were not eligible to hold land, and disempowered movers would have plausibly borne the brunt of forced labor.³⁵ Alternatively, low productivity individuals may have fled subjected villages.³⁶ 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

We begin by examining how the Cultivation System impacted land tenure in subjected villages, as there is an intensive focus on this question in the historical literature, starting with the seminal work *Agricultural Involution* by Clifford Geertz (1963) and emphasized in the most authoritative work on the Cultivation system written by Elson (1994). Village heads under the Cultivation System were empowered by the Dutch to distribute 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 (Elson, 1984, p. 94; Ricklefs, 2008, p. 159). Village land today is used for multiple purposes, including *tanah bengkok* - public land that the village head is allowed to use as compensation for his service.

We therefore begin by examining the allocation of land in more contemporary periods, focusing on village-owned land as well as 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 0.1 hectare land bins from the 2003 Agricultural Census.

The estimates show that the Cultivation System left a substantial mark on village-owned land. Table 2, columns (1) through (4) document 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). In 2003, about 1.5 percentage points more land was owned by the village in Cultivation System areas, relative to a sample mean of 9% of total land owned by villages. RD plots are shown in Appendix S.

Columns (5) through (8) examine 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.³⁷ Coefficients tend to be positive but fall short of statistical significance and are small compared to the mean.

³⁵While in theory landowners were responsible for cultivation services, in practice landless peasants attached themselves to landholding families and often performed the labor.

³⁶Elson (1984, p. 60) argues that the growth of frontier regions may have been due to individuals fleeing forced labor, though it is not obvious that the lowest productivity individuals would have been the movers.

³⁷Ownership data do not distinguish between agricultural and non-agricultural land, and hence we cannot examine inequality in agricultural land ownership.

5.2.2 Public goods

The village's ability to raise revenue through village owned land, as well as land tenure arrangements more generally, could impact human capital accumulation and the provision of local schooling (Acemoglu et al., 2009; Banerjee and Somanathan, 2007). This may have been particularly true prior to the massive INPRES school expansion, studied by Duflo (2001), as prior to INPRES village-level school construction was financed locally. This was true going back to the colonial period, when communities were responsible for raising a share of the resources if they wanted a village school (*volkschool*) (Aritonang, 1994).

Using the 100% sample of the 2000 Population Census, Table 3, 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 1.8 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 historically, we estimate effects by cohort using the 2000 Population Census. The left panel in Figure 13 plots estimates from equation (2). Each point represents a γ coefficient from a separate regression for a given cohort, with the 95% confidence interval indicated by a line. For ease of interpretation, the right panel plots sample means for each cohort and outcome variable. In general, schooling levels were initially very low and increased over time.

Impacts on years of schooling are large and positive across cohorts, even going as far back as the cohort born in the 1920s who completed their 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 3 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 13, panel g). One possibility is that the more generous compensation for village heads, in the form of more village land they can use ex-officio, attracted a relatively more educated pool of village heads. Empirical evidence from Indonesia in turn shows that more educated village heads increase public goods provision (Martinez-Bravo, 2017).

We also find positive effects on the presence of schools historically, using data on schools from the 1980 PODES village survey. Due to space constraints, these impacts are discussed in Appendix T. More generally, Appendix Table T-2 considers various public goods beyond education. Effects of forced cultivation tend to be positive but not statistically significant.

5.2.3 Economic structure

We finally turn to the overall structure of the economy. Higher human capital levels - and the disproportionately higher schooling levels of the village head - could plausibly lead to less agriculture and more manufacturing. Moreover, the historical literature 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).

Table 4, 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 in subjected villages, individuals are 4 percentage points (15 percent) less likely to work in agriculture, 3 percentage points (14 percent) more likely to work in manufacturing, and 1.2 percentage points (7 percent) more likely to work in retail.³⁸ Conditional on having agriculture, appendix U shows that Cultivation and non-Cultivation villages are growing similar crops. There does not appear to be a mechanical correlation between village land and the structure of the economy.³⁹

Columns (4) through (6) examine the structure of the economy in 1980, using data from the Population Census. The coefficient on employment in agriculture is similar in magnitude to that from the more modern SUSENAS data. 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 show a similar pattern. We find that subjected villages have around 22 percent more

³⁸The 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.

³⁹The correlation between the share of village-owned land and the share of population working in agriculture in Cultivation villages is only 0.05, whereas the correlation in non-Cultivation villages is -0.04.

manufacturing firms than non-subjected villages (column 7).

Finally, column (8) 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 differences in schooling.

5.3 Specification Robustness

Specification choices are inherent in an RD design, and we document that results are robust to alternative choices of: 1) RD bandwidth, 2) RD polynomial, 3) inclusion of additional controls, 4) sample inclusion criteria, and 5) additional features of the specification. Additional details are provided in appendices V through AF.

5.4 Discussion

The RD effects isolate the impacts of being subjected to forced cultivation and show that if anything, the persistent local development effects of forced cultivation reinforce the positive long-run economic impacts of creating a sugar processing infrastructure. This is surprising at first glance in the context of other studies that find negative development effects of colonial forced labor (Lowe and Montero, 2016; Dell, 2010), but there are 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), and impacts on public-use land persist through the present. While there are a variety of channels that could link the Cultivation System to long-run human capital accumulation and economic structure, the presence of village land is a particularly plausible mechanism, with revenue from village-owned lands facilitating public goods provision once the System was abolished and these lands returned to community use. More generally, what is notable here is the absence of the strong negative effects found in other settings.

6 Conclusion

This study documents that the establishment of a sugar processing infrastructure in colonial Java persistently increased industrialization, education, and consumption in areas near government sugar factories, even after the factories had disappeared. Similarly, villages forced to grow sugar cane for the Cultivation System have more schooling and manufacturing today. Infrastructure, agglomeration via input-output linkages, and human capital accumulation are important channels of persistence that plausibly reinforce each other.

These results highlight that the direction and magnitude of the long-run economic impacts of extractive institutions are inextricably linked with the specifics of how these institutions functioned, as these determine the channels of persistence. In the case of the Cultivation System, the positive impacts on economic activity plausibly dominated in the long run because of a constellation of features: processing had to be done on site, a modern transport infrastructure was built to connect the extensive network of sugar factories to international ports, and processed sugar - some of which was sold on local markets - had dense linkages to industries that remained important after Indonesian independence. In subjected villages, the higher share of village land plausibly raised the village's capacity to fund public goods, such as schools, once these lands were returned to community use.

The counterfactual is also important. In rural Java prior to 1830, landless peasants were customarily attached to landowning households through feudal-like norms, most production was for subsistence, the Javanese state was highly fractured, and financial markets were largely absent, all major and persistent barriers to industrialization that in 1830 did not show signs of disappearing. The Dutch made large-scale investments to overcome these barriers precisely because they could extract a large surplus. While it plausibly would have been better for the historical subjected populations - and potentially have led to better long-run outcomes - if the Dutch had attempted to industrialize Java through inclusive institutions that maximized the welfare of the Javanese populace rather than the profits of the Dutch treasury, this is not a feasible scenario given the objectives of the colonial state and indigenous power brokers.

The Dutch were not the only colonial power to reorganize colonial economies in order to create the surplus that they wished to extract, nor are the Javanese unique in productively utilizing colonial structures to generate economic activity long after the colonizers had left. These results plausibly inform a variety of contexts in which foreign powers attempted to extract a 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., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.
- ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

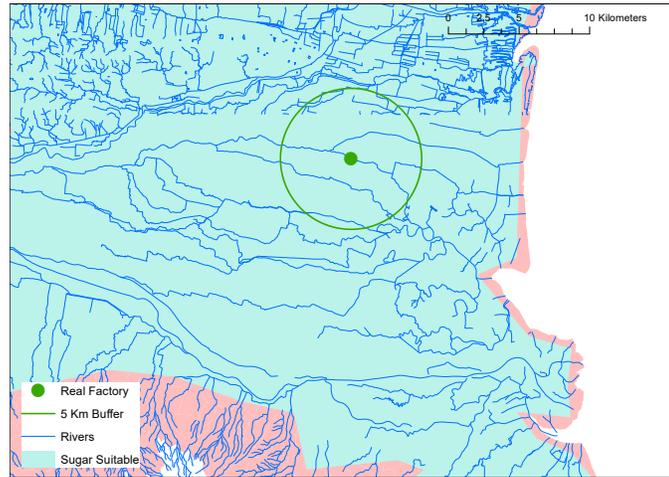
- (2002): “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution,” *The Quarterly Journal of Economics*, 117, 1231–1294.
- ANTLÖV, H. AND S. CEDERROTH (1994): *Leadership on Java: Gentle Hints, Authoritarian Rule*, Studies in Asian Topics. Richmond: Curzon Press/Nordic Institute of Asian Studies.
- ARITONANG, J. S. (1994): *Mission Schools in Batakland (Indonesia): 1861-1940*, vol. 10, Brill.
- BANERJEE, A. AND R. SOMANATHAN (2007): “The Political Economy of Public Goods: Some Evidence from India,” *Journal of Development Economics*, 82, 287–314.
- BAQAEE, D. (2015): “Labor Intensity in an Interconnected Economy,” Working paper, Harvard.
- BARTELME, D. AND Y. GORODNICHENKO (2015): “Linkages and Economic Eevelopment,” Tech. rep., National Bureau of Economic Research.
- BLEAKLEY, H. AND J. LIN (2012): “Portage and path dependence,” *The quarterly journal of economics*, 127, 587–644.
- BOOTH, A. (2007): *Colonial Legacies: Economic and Social Development in East and Southeast Asia*, University of Hawaii Press.
- BRAKMAN, S., H. GARRETSEN, AND M. SCHRAMM (2004): “The strategic bombing of German cities during World War II and its impact on city growth,” *Journal of Economic Geography*, 4, 201–218.
- CAGÉ, J. AND V. RUEDA (2016): “The long-term effects of the printing press in Sub-Saharan Africa,” *American Economic Journal: Applied Economics*, 8, 69–99.
- CAICEDO, F. V. (2014): “The mission: Human capital transmission, economic persistence and culture in south america,” Working paper, Bonn University.
- CARVALHO, V. M. (2009): “Aggregate Fluctuations and the Network Structure of Intersectoral Trade,” Working paper, Pompeu Fabra University.
- COEDÈS, G. (1966): *The Making of South East Asia*, University of California Press.
- COMMISSIE UMBGROVE (1858): “Archief van de Commissie voor de Opname van de Verschillende Suiker-fabrieken op Java [1853–1858],” Nationaal Archief, The Hague, Access Number 2.10.11.
- DAVIS, D. R. AND D. E. WEINSTEIN (2002): “Bones, bombs, and break points: the geography of economic activity,” *American Economic Review*, 92, 1269–1289.
- (2008): “A search for multiple equilibria in urban industrial structure,” *Journal of Regional Science*, 48, 29–65.
- DEATON, A. (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, World Bank.
- DELL, M. (2010): “The Persistent Effects of Peru’s Mining Mita,” *Econometrica*, 78, 1863–1903.
- DIPPEL, C., A. GREIF, AND D. TREFLER (2015): “The Rents from Trade and Coercive Institutions: Removing the Sugar Coating,” Tech. rep., National Bureau of Economic Research.
- DONALDSON, D. (2010): “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” Tech. rep., National Bureau of Economic Research.
- DONALDSON, D. AND R. HORNBECK (2016): “Railroads and American Economic Growth: A Market Access Approach,” *The Quarterly Journal of Economics*, qjw002.

- 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.
- ELLISON, G., E. L. GLAESER, AND W. R. KERR (2010): “What causes industry agglomeration? Evidence from coagglomeration patterns,” *American Economic Review*, 100, 1195–1213.
- 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.
- GALLEGO, F. A. AND R. WOODBERRY (2010): “Christian missionaries and education in former African colonies: How competition mattered,” *Journal of African Economies*, 19, 294–329.
- 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.
- HIRSCHMAN, A. O. (1960): *The Strategy of Economic Development*, Yale University Press.
- HUILLERY, E. (2009): “History matters: The long-term impact of colonial public investments in French West Africa,” *American economic journal: applied economics*, 1, 176–215.
- JEDWAB, R., E. KERBY, AND A. MORADI (2015): “History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya,” *The Economic Journal*.
- JEDWAB, R. AND A. MORADI (2016): “The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa,” *Review of economics and statistics*, 98, 268–284.
- KULKE, H. (1986): “Early and the Imperial Kingdoms in Southeast Asian History,” in *Southeast Asia in the 9th to 14th Century*, ed. by D. G. Marr and A. C. Milner, Institute of Southeast Asian Studies and Australian National University, Research School of Pacific Studies.
- LANE, N. (2017): “Manufacturing Revolutions: Industrial Policy and Networks in South Korea,” Tech. rep., IIES.
- LIEBERMAN, V. (1993): “Local Integration and Eurasian Analogies: Structuring Southeast Asian History, c. 1350-c. 1830,” *Modern Asian Studies*, 27, 475–572.
- 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>.
- (2017): “Mistrust in Medicine: The Legacy of Colonial Medical Campaigns in Central Africa,” Working paper, Harvard University.
- 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.
- MABBETT, I. W. (1977): “The ‘Indianization’ of Southeast Asia : Reflections on the Historical Sources,” *Journal of Southeast Asian Studies*, 8, 143–161.

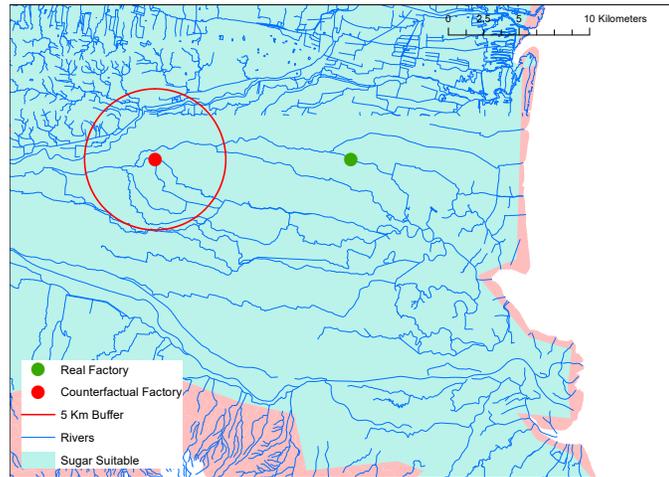
- MARSHALL, A. (1890): *Principles of Economics*, Macmillan and Co.
- MARTINEZ-BRAVO, M. (2017): “The Local Political Economy Effects of School Construction in Indonesia,” *American Economic Journal: Applied Economics*, 9, 256–289.
- MIGUEL, E. AND G. ROLAND (2011): “The long-run impact of bombing Vietnam,” *Journal of development Economics*, 96, 1–15.
- MOKYR, J. (1976): *Industrialization in the Low Countries: 1795-1850*, Yale.
- MYRDAL, G. (1957): *Economic Theory and Under-Developed Regions*, G. Duckworth.
- RASMUSSEN, P. N. (1956): *Studies in Inter-Sectoral Relations*, E. Harek.
- REDDING, S. J., D. M. STURM, AND N. WOLF (2011): “History and industry location: evidence from German airports,” *Review of Economics and Statistics*, 93, 814–831.
- 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.
- TAMBAH, S. J. (1977): “The Galactic Polity: The Structure of Traditional Kingdoms in Southeast Asia,” *Annals of the New York Academy of Sciences*, 293, 69–97.
- (2013): “The Galactic Polity in Southeast Asia,” *HAU: Journal of Ethnographic Theory*, 3, 503–534.
- 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 SCHAİK, 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.
- WOLTERS, O. (1999): *History, Culture and Region in Southeast Asian Perspectives*, Southeast Asia Program Publications, Cornell University.

Figure 1: Counterfactual Factories Example

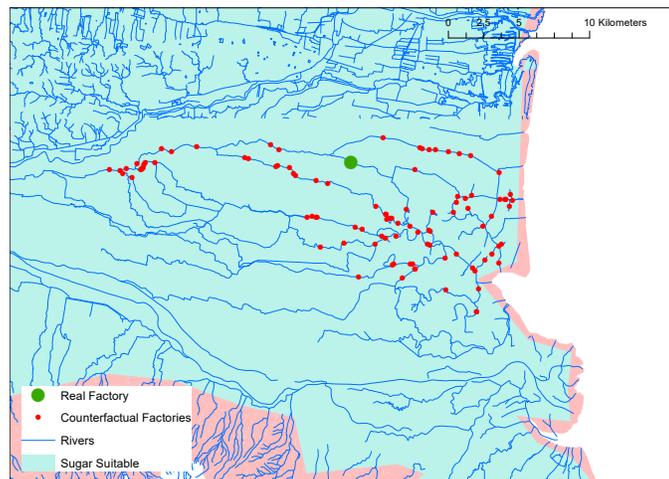
(a) Real Factory



(b) Counterfactual Factory Suitability

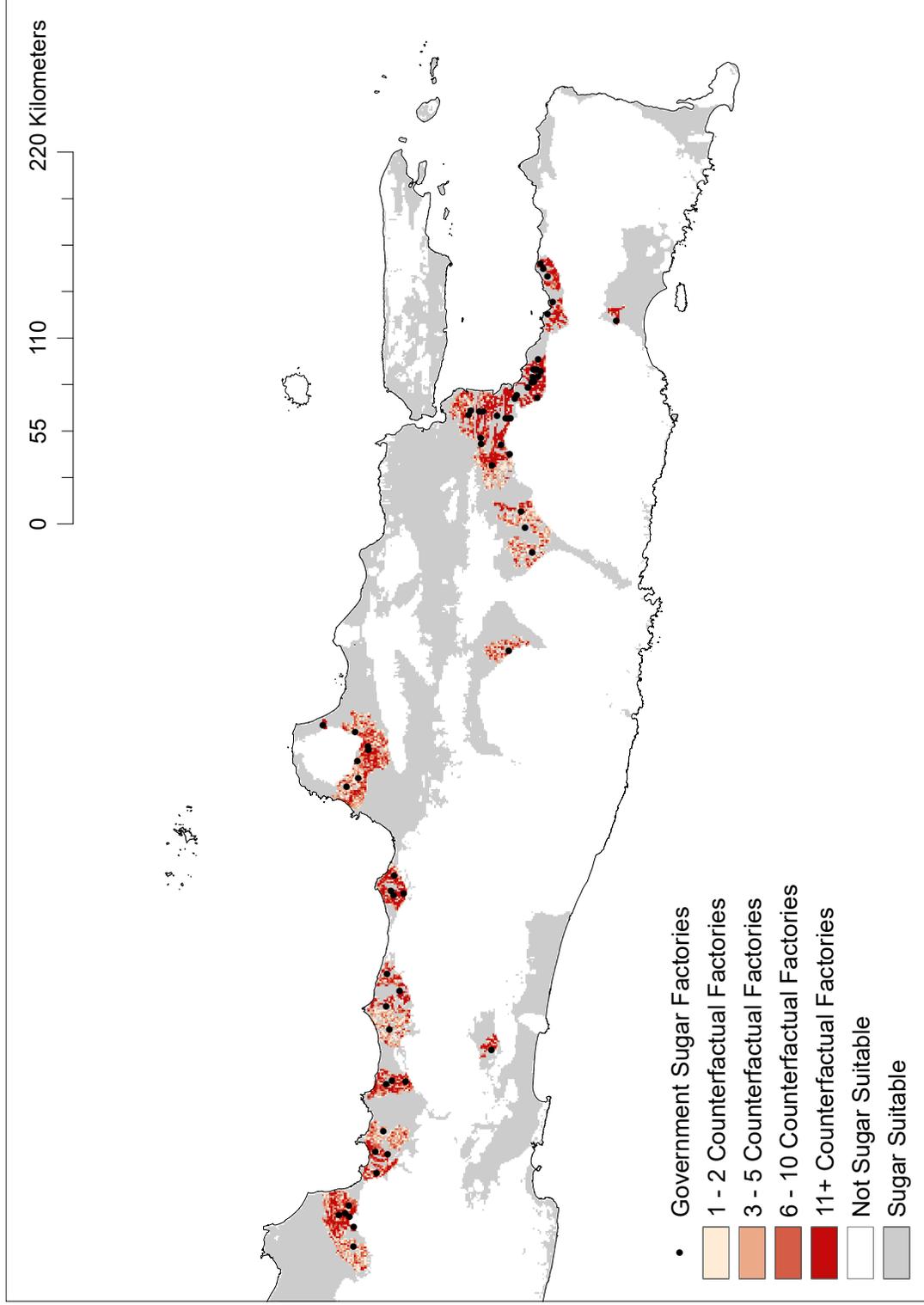


(c) Counterfactual Factories



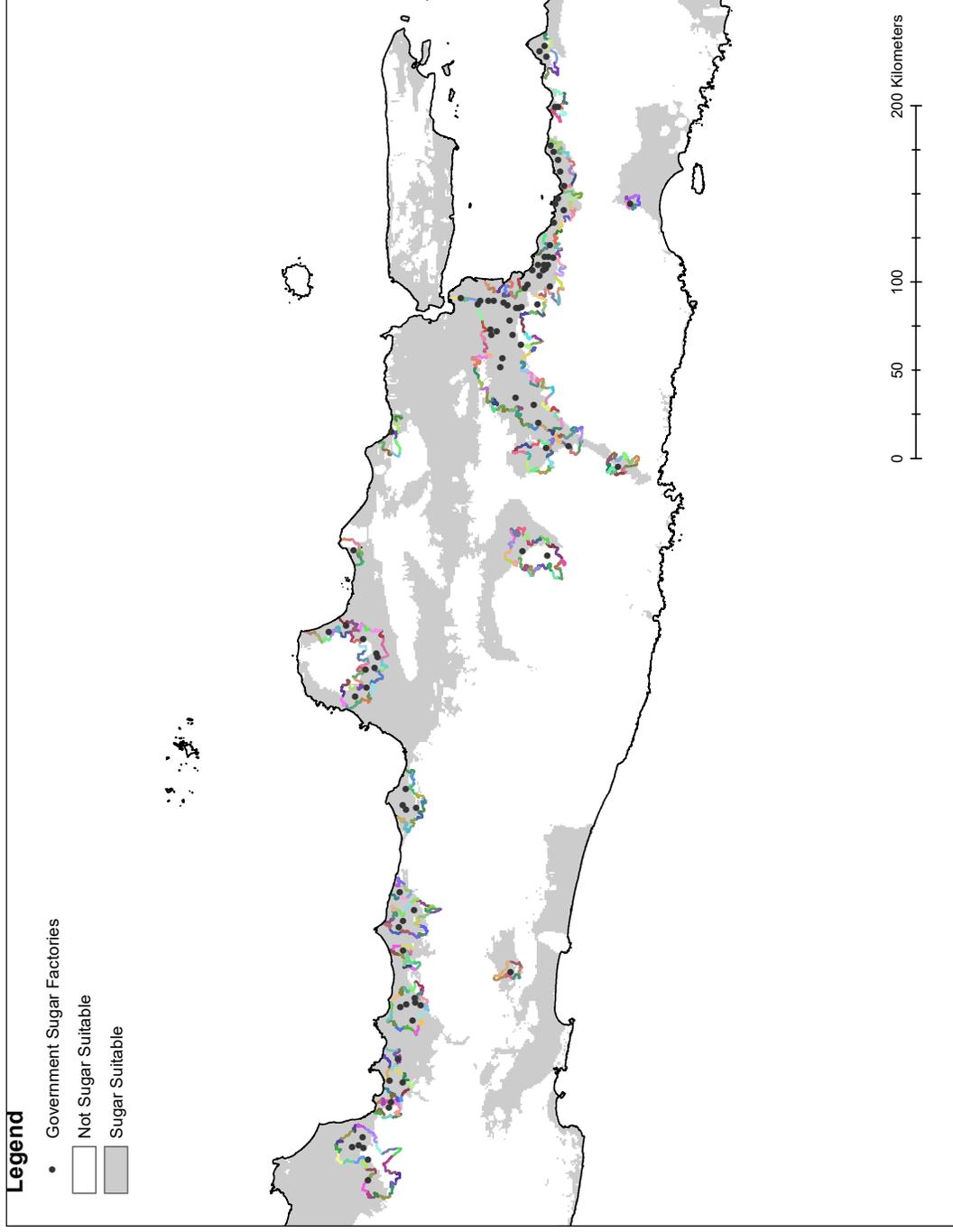
This figure illustrates the construction of the counterfactual factories, as described in Section 4.1.

Figure 2: Counterfactual Factories Heatmap



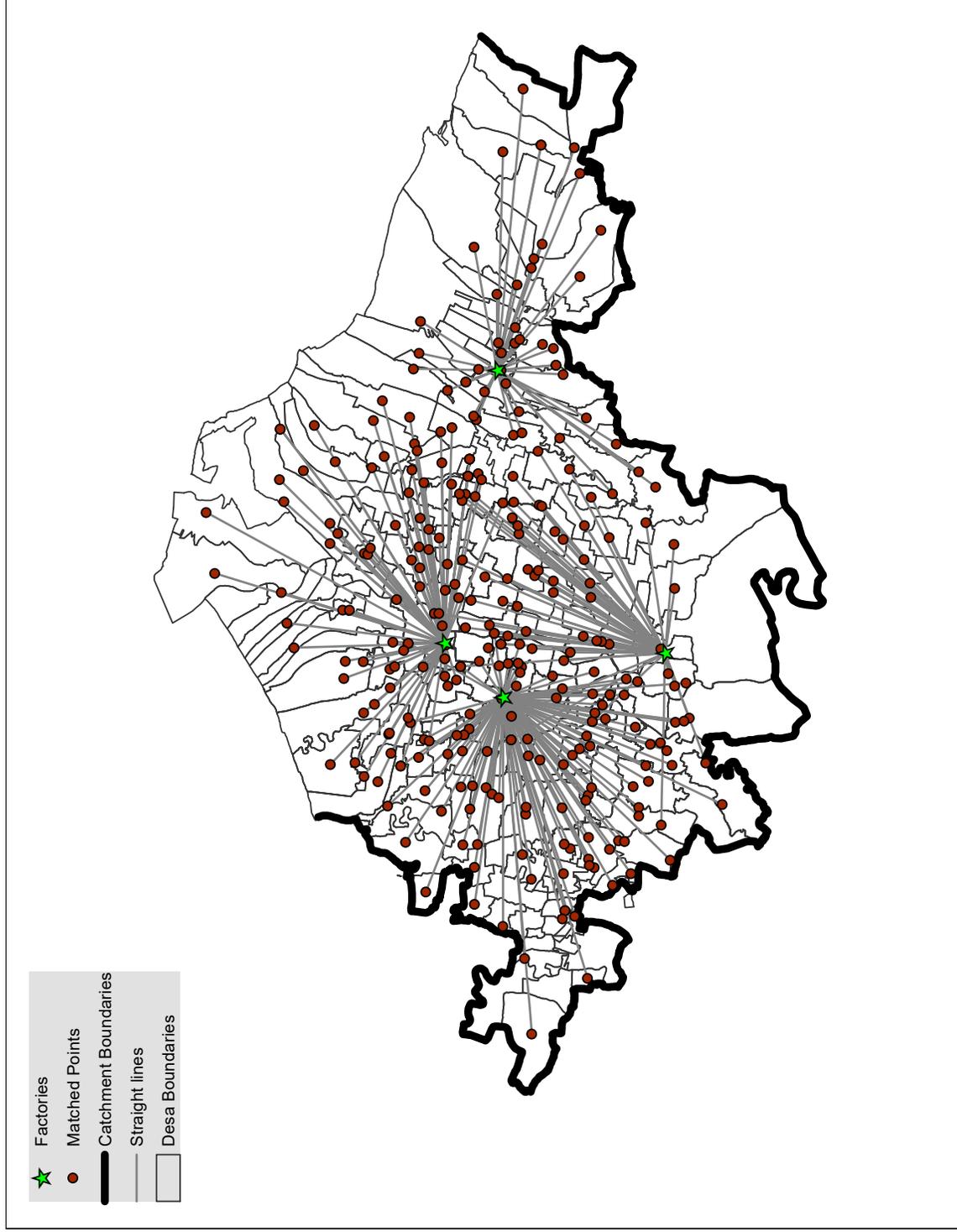
Black dots are the government sugar factories in the main sample. Gray areas have an elevation and slope that are suitable for sugar cultivation, whereas white areas do not. Within sugar-suitable areas, shading indicates the density of the counterfactual factories.

Figure 3: The Cultivation System



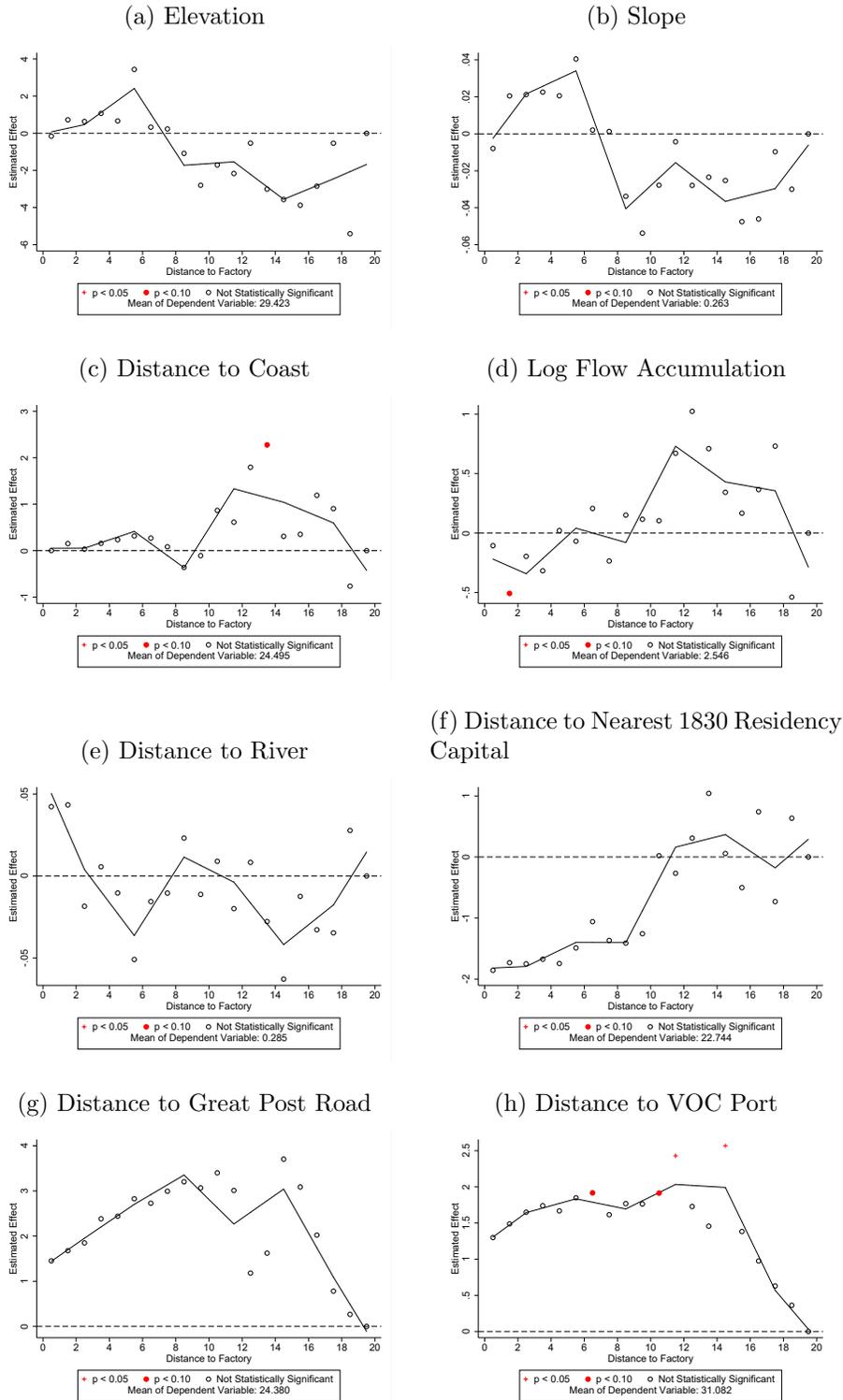
Sources: Commissie Umbgrove (1858). Black dots are government sugar factories. Lines are forced cultivation catchment area boundaries, with colors indicating 10 km boundary segment fixed effects. Gray areas have an elevation and slope that are suitable for sugar cultivation, whereas white areas do not.

Figure 4: Construction of Catchment Areas



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

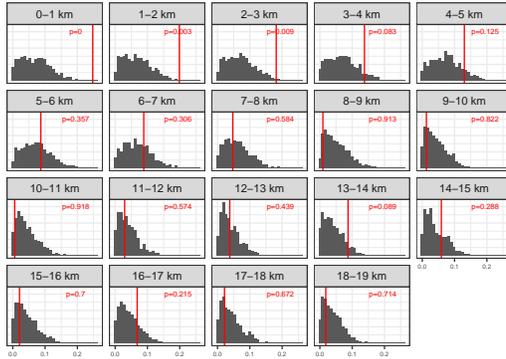
Figure 5: Geography



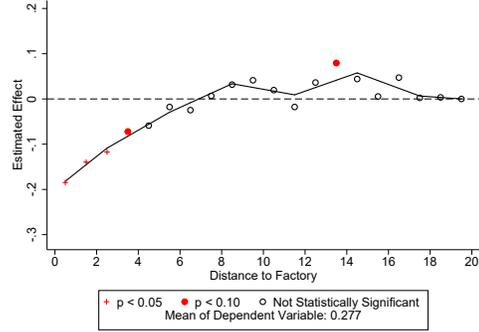
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure 6: Share in Agriculture (2001-11): Illustration of Methodology

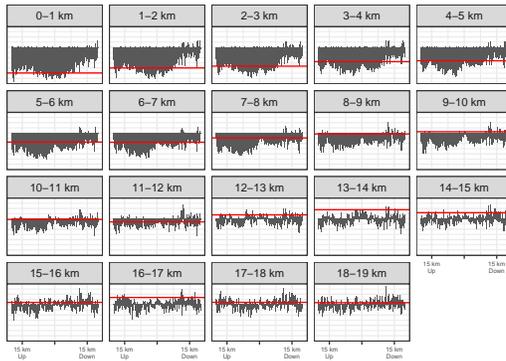
(a) Independent Shifts: Counterfactuals



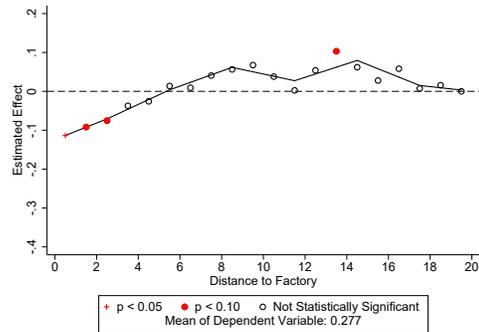
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals

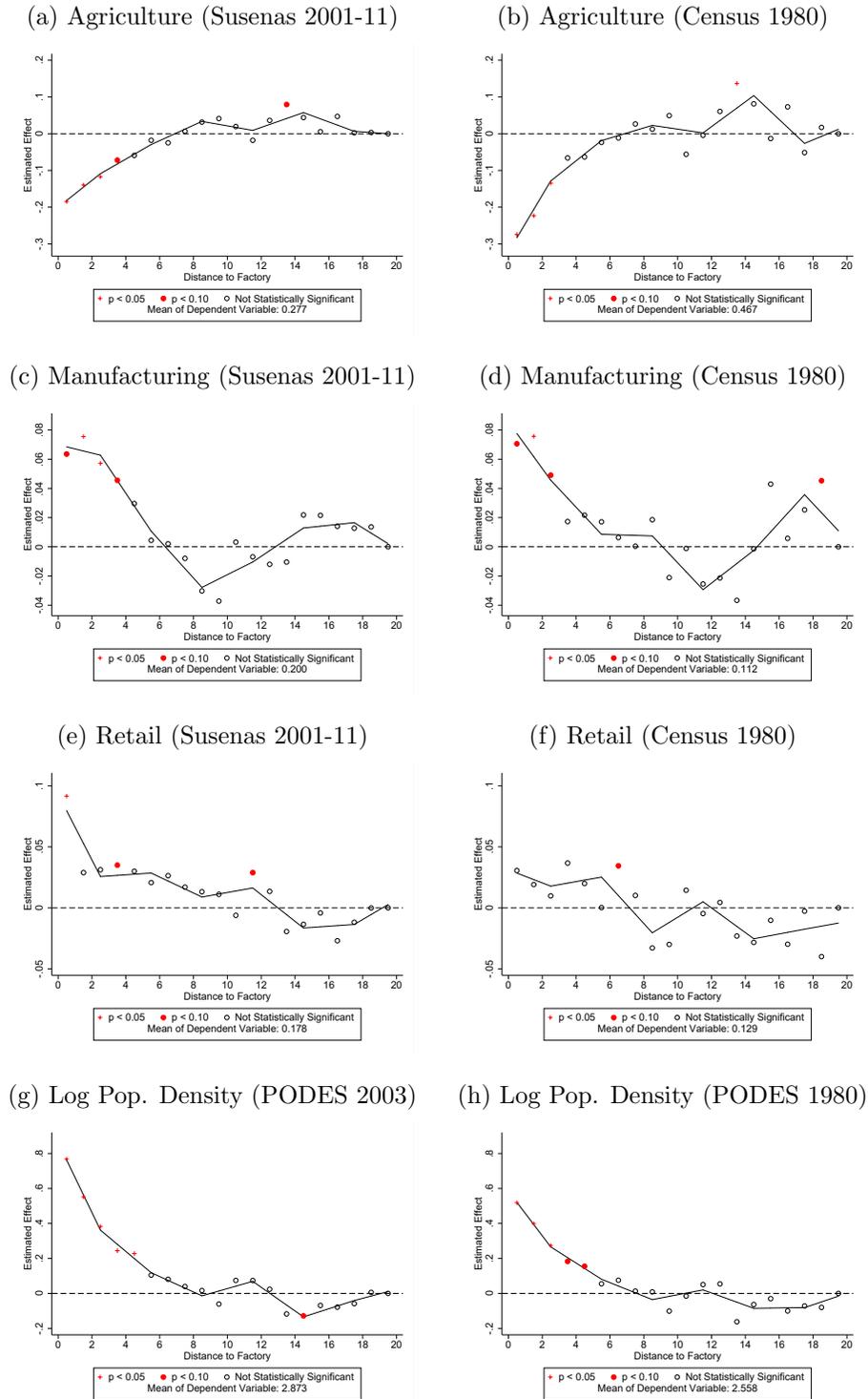


(d) Common Shifts: Plotted Coefficients



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 and pre-period characteristics, 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 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions 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. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

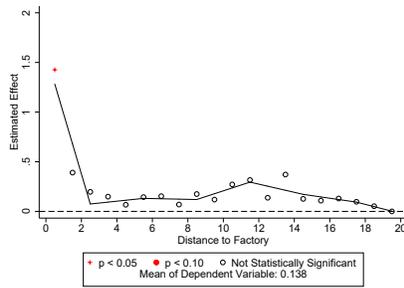
Figure 7: Industry and Agglomeration



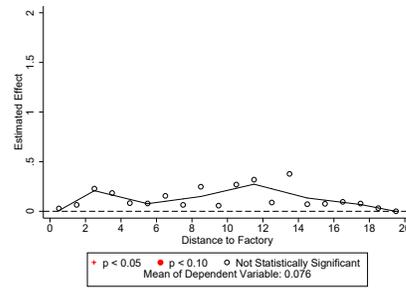
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. 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 points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure 8: Sugar and Linked Industries

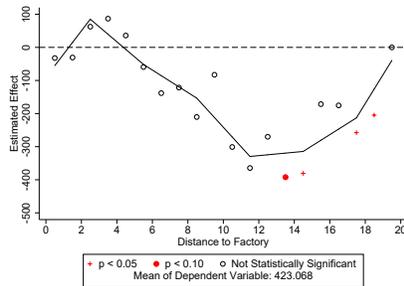
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



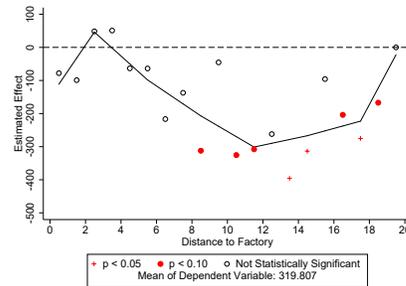
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



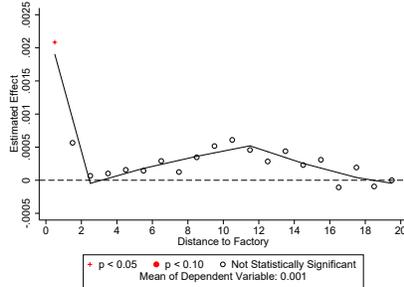
(c) Tons of Cane Grown (Full Sample, PODES 2003)



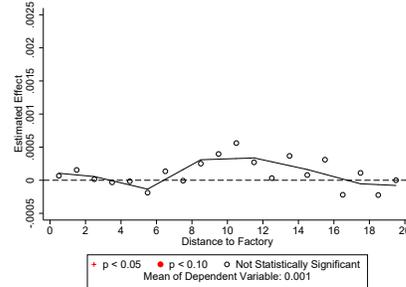
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



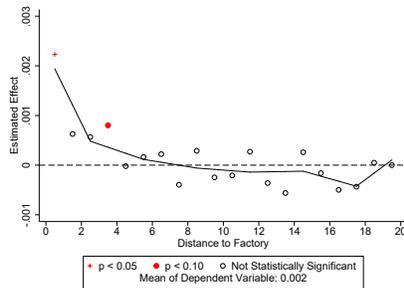
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



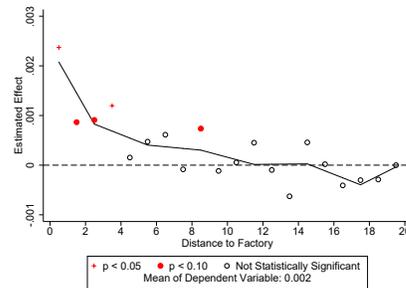
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



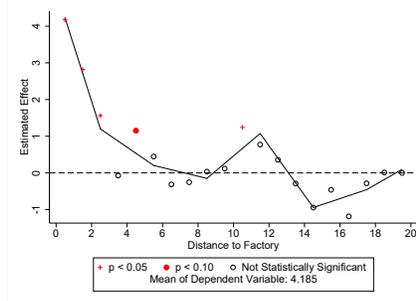
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



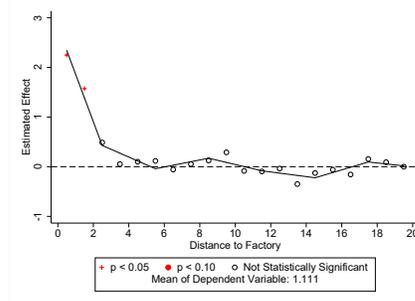
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure 9: Infrastructure

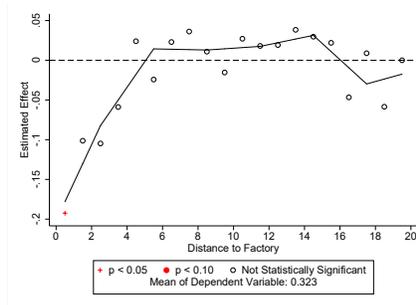
(a) Colonial Road Density (1900)



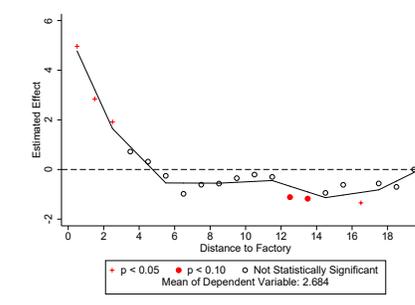
(b) Colonial Railroad Density (1900)



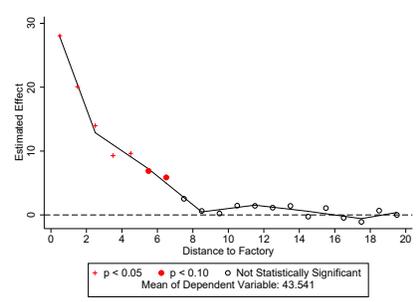
(c) Dirt Road (PODES 1980)



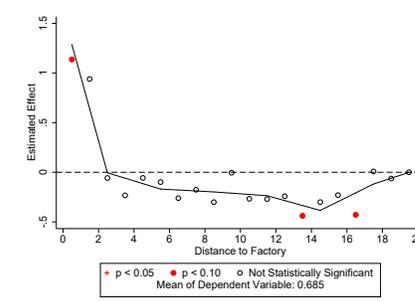
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



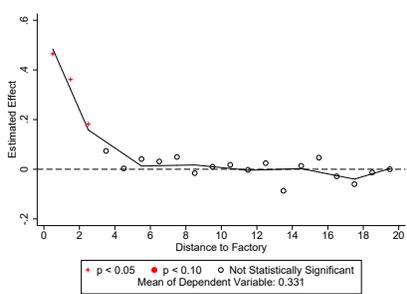
(f) Railroad Density (2017)



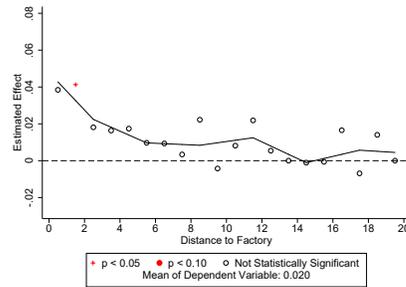
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure 10: Other Public Goods

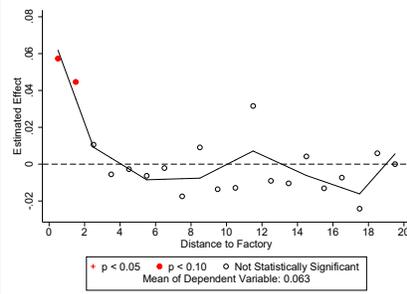
(a) Village Has Electricity (PODES 1980)



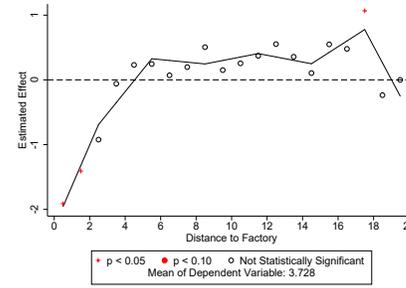
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)

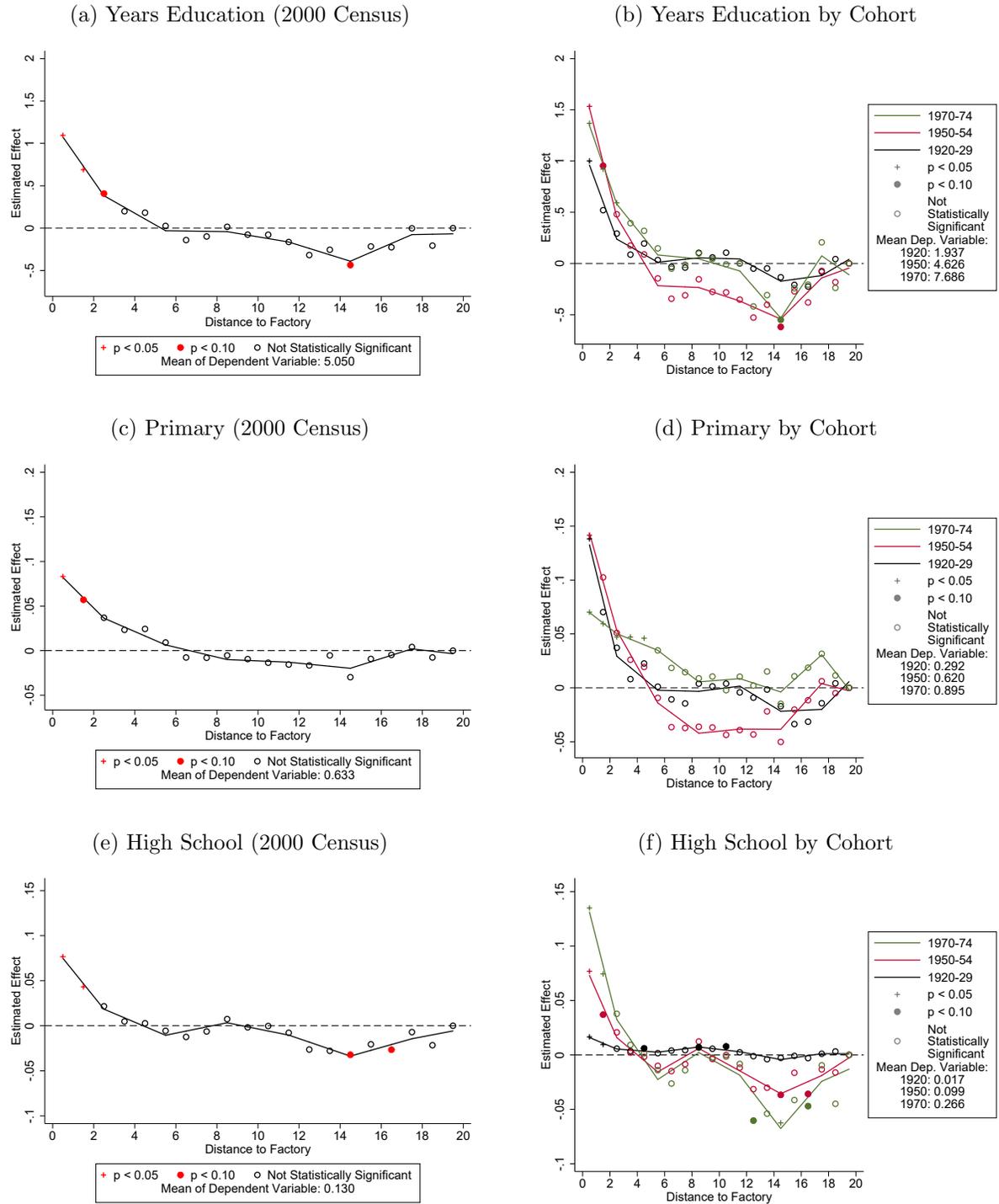


(d) Distance to Subdistrict Capital (2011 PODES)



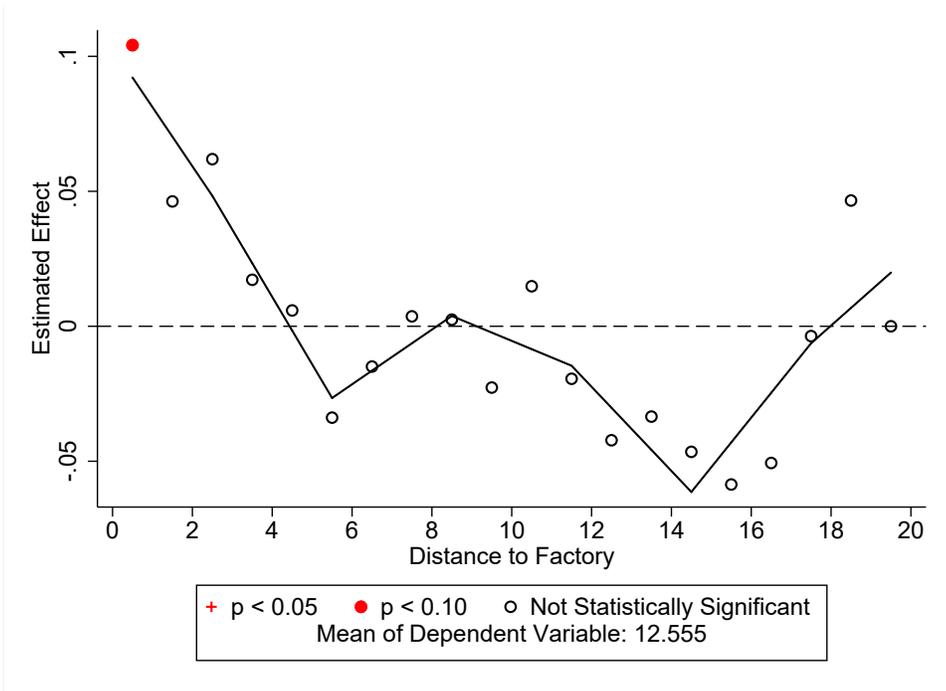
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure 11: 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, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

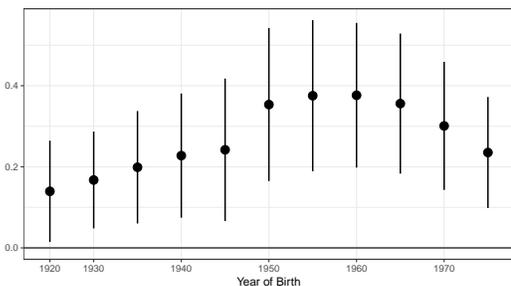
Figure 12: Expenditure



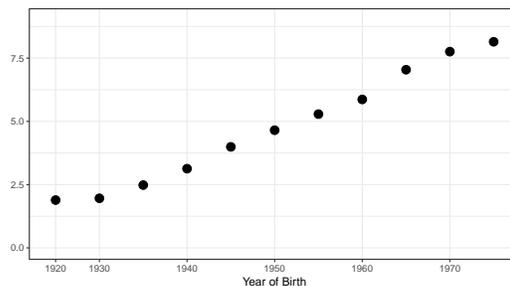
Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure 13: Education by Cohort: Subjected Villages (2000 Census)

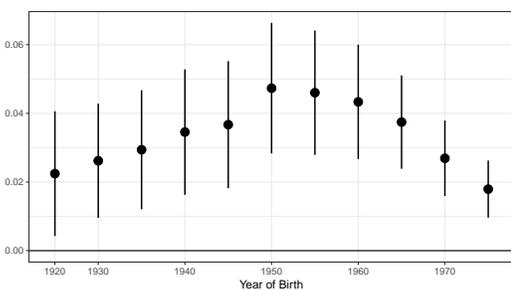
(a) Years of Schooling



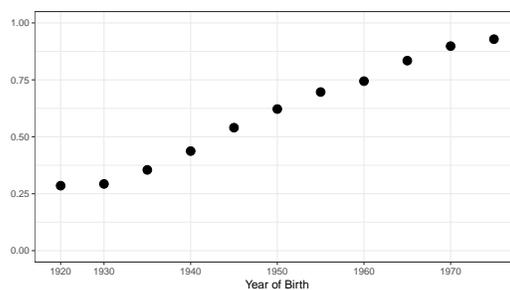
(b) Years of Schooling Levels



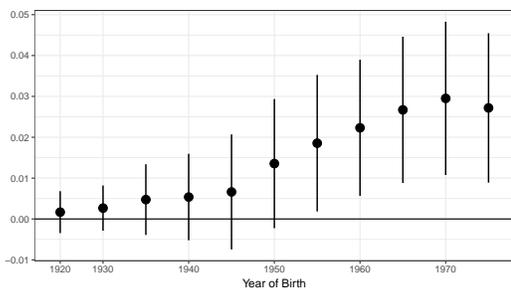
(c) Primary Completion



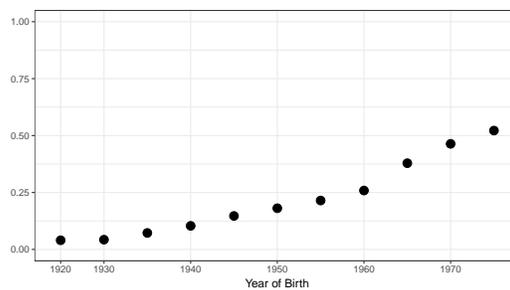
(d) Primary Completion Levels



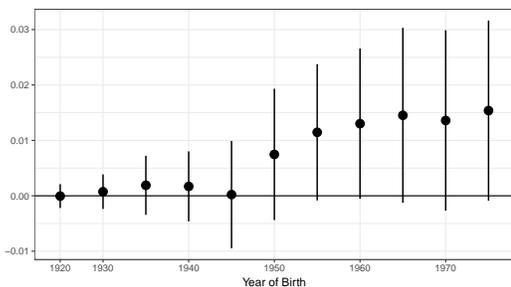
(e) Junior High Completion



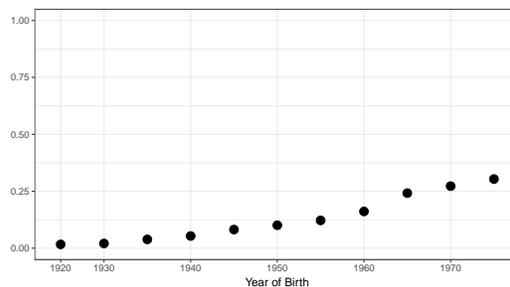
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



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 and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, 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	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-1.908 (0.744)	-0.017 (0.013)	0.117 (0.106)	-0.041 (0.160)	0.001 (0.012)	0.062 (0.133)	-0.060 (0.118)	0.265 (0.136)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

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		Village Land 1980		99th Pctile	90th Pctile	90th Pctile	50th Pctile
	Total Land	Land Share	Total Land	Land Share	÷ 90th Pctile	÷ 10th Pctile	÷ 50th Pctile	÷ 10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.339 (0.851)	0.015 (0.004)	3.631 (1.559)	0.013 (0.005)	0.032 (0.150)	0.860 (1.130)	0.215 (0.191)	-0.046 (0.141)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, 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: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.220 (0.074)	0.025 (0.006)	0.015 (0.007)	0.007 (0.005)	-0.018 (0.008)	-0.004 (0.010)	0.094 (0.086)	0.028 (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 boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, 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 4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf. Firms	Log Equiv. Consumption
	Ag. (1)	Manuf. (2)	Retail (3)	Ag. (4)	Manuf. (5)	Commerce (6)	(7)	(8)
Cultivation	-0.042 (0.012)	0.031 (0.010)	0.012 (0.007)	-0.036 (0.021)	0.008 (0.013)	0.027 (0.010)	21.734 (8.639)	0.006 (0.011)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.