Spillovers from agricultural processing

Ryan B. Edwards[*]
Australian National University

July 2, 2024

Abstract

This paper uses the proliferation of palm oil factories across Indonesia's undeveloped hinterland to study industrial onset and estimate spillovers from agricultural processing. The main finding is signs of urbanization and structural change around factories: more non-agricultural employment, higher incomes, and more people, firms, and other economic and social organizations. These patterns are largely explained by economic linkages, infrastructure and other public goods, and economies of scale in production. By focusing on subsistence rural regions in a large developing economy, this paper adds a globally-significant new case to a growing literature emphasizing the importance of agglomeration externalities for understanding the birth of new towns, the spatial distribution of economic activity, and structural transformation.

JEL codes: F14; F23; F63; J43; O13; O14; O19; O53; Q17; R11

Keywords: agricultural processing; urbanization; structural change; economic organisation; local labour markets; Indonesia; palm oil

[*]Development Policy Centre and the Arndt-Corden Department of Economics, Crawford School of Public Policy, Australian National University. Contact address: 7 Liversidge Street, Australian National University, Canberra ACT 2601. Email: ryan.edwards@anu.edu.au. Thanks to Treb Allen, Chris Barrett, Sam Bazzi, Marc Bellemare, Amy Damon, Taryn Dinkelman, Eric Edmonds, David Eil, Wally Falcon, Jim Feyrer, Gordon Hanson, Paul Goldsmith-Pinkham, Will Johnson, Amy Lastuka, Blane Lewis, Paul Novosad, Emilia Tjernstrom, Dietz Vollrath, two helpful anonymous referees, and seminar participants at ASEAN, ANU, Dartmouth College, Middlebury College, the University of Michigan, and the World Bank for helpful comments and discussions. Grace Hadiwidjaja, Sebastian Kraus, Budy Resosudarmo, Mayang Rizky, and Pablo Pacheco kindly shared data. Lauren Archer, Robert Castle, and Victoria Meyer provided excellent research assistance under the Dartmouth Presidential Scholars program.

1 Introduction

Can export-oriented agricultural processing help ease the transition out of agricultural employment? Transitioning the majority of workers in developing countries to higher-value, market-based agricultural activities and ultimately out of agriculture is essential for breaking rural poverty traps (Gollin, Lagakos, and Waugh, 2014). Characterized by agro-industrial conglomerates sourcing from both industrial and small, family farms, modern agricultural value chains reach from cities like Rotterdam, Saint Louis, Seattle, and Singapore to some of the most remote parts of the developing tropics (Byerlee, Falcon, and Naylor, 2016; Bellemare and Bloem, 2018; Barrett, Reardon, Swinnen, and Zilberman, 2022). By imposing industrialization on subsistence agrarian communities, new export-oriented processors could dramatically reorganize rural economic life for better or for worse: offering a foothold into global markets, value chains, and industrialization, or leading to two-tiered labor markets, a loss of land among the poor, and immiserisation of local communities. This paper uses the proliferation of palm oil factories across Indonesia's remote outer islands since the late 1990s to study the onset of industry, test these competing theories, and estimate the local economic impacts and spillovers to non-agricultural activity from agricultural processing. In doing so, I shine new light on and quantify the various channels through which new industrial activity affects local labour markets in rural parts of developing countries.

The main empirical concern is purposive placement: communities that receive a factory may already differ from those that do not. An ideal natural experiment might see factories arbitrarily scattered across an undeveloped hinterland, or following some observable placement rule. Several unique features of the palm oil supply chain and Indonesia's expansion are helpful in this regard. Indonesia's fourfold increase in palm oil production since the late 1990s is one of the world's largest modern agricultural expansions (Byerlee, Falcon, and Naylor, 2016). Since most Indonesian palm oil is exported, the relevant demand is external to producing communities (Gaskell, 2015). On the supply side, expansion into unindustrialized rural frontiers helps reduce concerns about firms colocating to capture agglomeration economies from other firms.

Figure 1 uses three satellite images to illustrate how things typically unfold. The first image was taken in 1984, when one of Indonesia's largest firms started developing plantations in the Kerinci area of Riau province, the largest producing region today. The second picture was taken in 2000, as the current expansion commenced. The third image shows the urban center today, surrounded by a mosaic of commercial estates and small, family farms.[1] Since oil palm fruits must be processed within 24 hours, direct impacts tend to be concentrated near factories and factory placement closely follows observable growing conditions. Specifically, communities that happen to be in areas most suitable for palm cultivation experienced growth in processing because the optimal factory location is the point that minimizes the amount of land needed to feed the factory while maximizing feed quality.

My empirical strategy exploits spatial differences in factory exposure. I create a new geospatial dataset collecting the universe of Indonesia's palm oil processors—one of the world's leading drivers of tropical deforestation (Hsiao, 2024)—and compare villages near factories to similar but less exposed villages slightly farther away.[2] Treatment effects are relative within localities, and this between-village variation offers a natural long-run interpretation appropriate for studying the roots of urbanization and structural change. A rich control function captures relative suitability within factory catchments based on precisely measured topographic, hydrological, and other geographic characteristics, while characteristics unrelated to suitability tend to be balanced near and farther from factories. To address potential reverse causality concerns, I show that there are no statistically significant effects on pre-period population, infrastructure, industry, and other outcomes, consistent with new factories moving into the hinterland and the time dynamics shown in Figure 1. I additionally show that the results are quantitatively similar when including a demanding set of additional control variables, and present

[1]Figure A20 provides some "street view" images taken from Google Maps. Kerinci's transformation is not dissimilar to more nascent areas since 2000, and similar time series images from Kalimantan and Papua are provided in Figures A21 and A22, and a close-up of recent developments in Bengkalis, Riau in Figure A23.

[2]These simple and transparent spatial difference comparisons are closely related to and build on the approaches developed in Dell and Olken (2020) and Fafchamps et al (2016).

consonant estimates from multiple instrumental variables (IV) strategies and falsification tests exploiting the reduced form. Together, these exercises suggest that the main results are likely due to new factories rather than some unobserved variable or preexisting condition.

The main finding is signs of urbanization and structural change around palm oil factories. Using the complete census micro-data, I document that people living closer to factories are eight percent more likely to work in non-agricultural sectors (i.e., industry and services). People living near factories are almost twenty percent more likely to be employed, rather than self-employed or working at home. Per capita household expenditures are ten percent higher near factories. Agricultural households appear to drive the consumption gains, although magnitudes are similar across sectors. Together, these local labor market estimates paint a picture of villages around palm oil factories as embryonic towns undergoing a process dubbed proto-urbanization by Marcel Fafchamps, Michael Koelle, and Forhad Shilpi (2016). Higher population and firm densities around factories confirm this interpretation. The average village 5–10 km from a palm oil factory has 500 more people and one more firm (excluding palm oil processors). Effects are twice as large within 5 km. All labor market results are similar for recent migrants and locals, suggesting employment effects are not simply a mechanical effect of population growth. There appear to be significant local benefits from agricultural processing beyond those accruing to farmers.

How could relatively unsophisticated processing facilities lead to such a dramatic reorganization of rural economic activity? Linked industries are the first potential explanation. Villages near factories tend to have more firms and organizations of many different types, including those that provide inputs to the palm oil sector. However, in contrast to Dell and Olken (2019), which highlights the importance of linkages for the persistent effects of Javanese sugar processing, effects here are approximately four times as large for un-linked industries. Neighboring firms range from micro-enterprises to large firms and are mostly in the retail and maintenance, finance and insurance, transport and warehouse, construction, and processing sectors. Villages near factories also have more cooperatives and banks. Although I find some evidence of backward linkages, the patterns are more consistent with demand-based consumption linkages and broader agglomeration forces.

Thus, a second explanation may be related to infrastructure and publicly-provided goods. Export-oriented agricultural manufacturing often requires up-front investments, not only in factories but also in transport and utilities networks. I find that villages around factories are indeed more likely to have (a) the main road upgraded from dirt to gravel or asphalt (within 20km), (b) public transport (within 5km), (c) lighting on the main street (within 50km), and (d) fewer households without electricity (within 30km). Improved infrastructure, local market integration, and market access are likely to reinforce any shift out of subsistence food production towards market-oriented agriculture and non-agricultural employment.

A booming economy also provides apt opportunities for local governments to raise revenue and provide more public goods, which could generate higher returns with industrial production nearby. Villages' annual budgets are, on average, about 50 percent larger within 10km of a factory. Villages near factories also rely less on inter-governmental transfers, have higher expenditures, and own more land, buildings, and other assets. Publicly-provided goods not closely related to agricultural supply chains are more common, specifically marketplaces, schools, health facilities, and places of worship (e.g. mosques and churches). To gauge whether village fiscal windfalls are driving the increase in public goods, I adjust estimates for local government revenue and expenditure. Results are statistically indistinguishable, suggesting that at least some of the new public goods near factories may be privately provided or due to targeted cash or in-kind inter-governmental transfers (i.e., not from own-source revenue).

Improved infrastructure could decrease trade costs, opening up the possibility of scale economies in production as a final channel at work. For each village, I calculate the number of potential buyers of palm oil fruits as a proxy for farm-to-factory market density. I find that economic agglomerations are strongest for the subset of villages near factories with other factories also within marketing distance. After the first factory opens up a local agricultural export market, thicker farm-to-factory markets appear to reinforce agglomerations further.

This study contributes to three streams of economics. I first extend the classic literature on agriculture, industrialization, and economic development, particularly that focused on linkages from agriculture to industrial and service sectors (Clark, 1940; Myrdal, 1957; Hirschman, 1960; Johnson and Mellor, 1961; Schultz, 1964; Chenery and Syrquin, 1975; Marden, 2018). Emphasizing consumption linkages from rising agricultural incomes and production linkages from locally-sourced inputs for processed agricultural products, this work is predicated on the idea that agricultural processing offers an entry point to better jobs with limited skill requirements. Despite a rich intellectual history and continued emphasis from policy-makers, empirical support for these theories remains thin, particularly in relation to modern global agricultural value chains.[3]

At least four recent studies make headway. For example, Melissa Dell and Benjamin Olken (2019) study the Dutch Cultivation System on Java. Technology constraints at the time saw sugar factories placed along rivers, adjacent to enough sugar-suitable land, and a reasonable distance apart. Using a selection-on-observables design similar to this paper, they find persistent positive impacts from sugar processing. Paula Bustos, Bruno Caprettini, and Jacopo Ponticelli (2016) study the more recent Brazilian soy expansion. Soy is also exported and requires local processing. Comparing sectoral employment, wages, and productivity across regions, they find that Brazil's soy expansion led to local structural change. In a companion paper, I compare regional poverty and consumption trajectories over Indonesia's expansion to find broad consumption gains in producing districts (Edwards, 2019). Christoph Kubitza and Esther Gerhke (2018) find fertility reductions with a similar approach. Here, I extend the scope of Dell and Olken's analysis to the world's largest modern agricultural expansion and a contemporary setting of decentralized and relatively nascent democratic institutions. I extend the other three studies to a much finer level of spatial aggregation with a focus on processing and clarifying some underlying channels at work.

[3] Gollin (2010), Dercon and Gollin (2014), and Bellemare and Bloem (2018) provide surveys.

My study also relates to work at the nexus of trade, spatial development, and economic geography.[4] Normally, factory-led industrialization happens in growing areas with good market linkages and rising education. It is not clear whether factory-led development can succeed in places without these characteristics. Indonesia's palm oil expansion offers the ideal empirical setting to learn about how industry affects local labor markets in developing countries, a unique natural experiment to study the impacts of new factories in places that might be viewed as unprepared to industrialize. Far from adverse effects, I find that people adapt quickly, benefits are broad-based, and consumption linkages, infrastructure, and scale economies reinforce agglomerations. This study thus adds a unique and globally significant new case—one of the first focused on largely subsistence rural regions in a large developing economy—to a growing literature emphasizing the importance of agglomeration externalities for understanding the birth of new towns, the spatial distribution of economic activity, and structural transformation.

Finally, my findings inform a topical debate on sourcing from developing countries (Swinnen, 2007; Maertens and Swinnen, 2009; Dragusanu et al, 2014). The traditional view is that large commercial farms are not only harmful for the environment but also socioeconomic development in producing communities, despite their increasing connectedness to smallholder systems through contract farming and other modalities (Byerlee, de Janvry, and Sadoulet, 2009; Easterly, 2007; Engerman and Solokoff, 2002; Farina and Reardon, 2000; World Bank, 2008). While these concerns are not limited to palm oil, controversy over palm oil has been so prominent the Indonesian President banned new permits in September 2018 and the European Parliament imports in 2017. The World Bank has had a moratorium in place since 2009. An alternative view is that agricultural processing might provide a foothold on the ladder of industrial development, offering capital inflows and market access for areas typically lacking both. My findings echo those of Dell and Olken (2019) and Bustos et al (2016), more consistent with the second view.

[4]See, e.g., Michaels, Rauch, and Redding (2012), Allen and Arkolakis, 2014), and Donaldson (2018). Fujita, Krugman, and Venebles (1998), Moretti (2010), and Donaldson (2015) provide more general surveys from the perspectives of new economic geography, local labor markets, and market integration.

The next section provides a brief introduction to the Indonesian palm oil sector and explains factory placement. Section 3 details the empirical strategy. Section 4 presents the main results on local labor markets and rural agglomeration. Section 5 explores economic linkages, rural infrastructure and other public goods, and economies of scale in palm processing as potential explanations for the main findings. Section 6 concludes.

2 Palm oil processing in Indonesia

Palm oil is the world's most consumed vegetable oil, produced almost entirely in tropical developing countries. Oil palms yield more oil per hectare than any other crop, helping to manage growing pressure on farmland from more land-intensive oils like soy and rapeseed (Corley and Tinker, 2016). Global palm oil production has roughly doubled every decade since the 1960s, increasing from less than 5 million metric tonnes per year in 1970 to over 70 million in 2015 (Byerlee et al 2016). Indonesia supplied 55 per cent of the 65 million metric tons produced globally in 2016–17. Spurred by sweeping decentralization reforms and a devalued rupiah following the Asian financial crisis (Rada, Buccola, and Fuglie, 2010), the area under cultivation for oil palm in Indonesia increased from 2.9 million hectares in 1997 to over 12.5 million today (Figure A24). Decentralization ushered in a parallel regime shift in production from select, centrally-planned plantation developments to liberalized investment and diffuse growth encouraged by newly-empowered local governments (Fitriani, Hofman, and Kaiser, 2005; Naylor, Higgins, Edwards, and Falcon, 2019). Since oil palm fruit must be processed within 24 hours to be of a high enough quality for global markets, cultivation is concentrated around factories. A new factory effectively opens up a new export market for local farmers, and smallholder oil palm adoption around new factories tends to be voracious. From 2008 to 2015 alone the number of palm oil businesses (most of which are processors) increased by 43%, from 1,059 to 1,511 (BPS, 2017). Each factory relies on dozens of surrounding villages, usually in addition to a primary plantation owned by the same firm, for steady supply.

I trace the supply chain from the farm gate through the factory to final export markets in Figure 2. Upstream, smallholders manage over 40% of 12 million hectares planted with oil palm. The remaining 60% is large industrial estates, around 20 per cent of which is estimated to have displaced forest (Austin et al., 2019). Road networks and transport logistics are needed to immediately transport fresh harvests to factories for timely processing. In undeveloped rural areas without prior industry, firms or governments typically need to build the infrastructure. With logistics networks in place, smallholders then aggregate and coordinate their activities through a complex network of cooperatives, traders, and other aggregation points (e.g., village loading ramps). Individual farmers typically report higher incomes and labor-saved after entering the sector (Krishna et al 2017; Kubitza and Gerhke, 2018).[5] Each factory also employs around 200 highly-skilled workers which, together with smallholder adoption, could generate consumption linkages through demand for locally produced goods and services.

Indonesia's palm oil factories sell mostly to refineries and global commodity markets. According to the 2010 Input-Output Table, nine domestic industries use palm oil and byproducts as inputs: animal and vegetable oil, peeled grain, animal feed, pesticide, drugs, pharmaceutical products, soap and cleaning, industrial cosmetic products, and other chemical goods industries (BPS, 2010). On the input side, the palm oil industry draws from 60 out of 192 sectors (see Appendix B for the complete listing). Hence, any economically important production linkages are more likely to be on the input side.

2.1 Factory placement

Understanding why factories are located in particular places is crucial for identifying their impacts. Despite growing smallholder involvement in the post-Suharto era, most palm oil factories still operate as part of a plantation system rather than as stand-alone entities. There are three key requirements to develop plantation systems: (a) final product demand, (b)

[5]Note that this labor-saving characterization only really applies on a per hectare basis or for a smallholder household holding plot size constant. At the regional level, any labour saved on a per-hectare or individual basis has generally been offset by the additional area under cultivation (Edwards, 2024).

financial capital to build the factory, related supply chain infrastructure, and nearby plantation (traditionally known as the "nucleus" estate in Indonesia), and (c) sufficient suitable cultivation area nearby to operate the factory efficiently (Hayami, 2010; Pryor, 1982).

As an export-oriented commodity experiencing an unprecedented surge in global demand (over a 300% increase since the 1990s), new factories arise principally due to external rather than local demand conditions. Large up-front costs see investments in processing infrastructure made by large firms based in capital cities or abroad (c.f., locally). Palm oil expansion has been across Indonesia's relatively undeveloped outer islands. Investments are usually "greenfield", introducing industry to subsistence agrarian communities. The need for immediate processing means direct impacts are then concentrated close by. These conditions alone set up an ideal natural experiment to study the onset of industrialization and spillovers from new factories. By contrast, the impacts of any one factory are usually difficult to disentangle because factories tend to locate near existing economic activity to capture agglomeration economies (Greenstone, Hornbeck, and Moretti, 2010).

Whether one relatively undeveloped rural community gets a factory is typically decided by the central government and district heads (i.e., bupatis). A burgeoning qualitative literature documents how—under the New Order and in the post-Suharto "reformasi" era—plantation and natural resource projects tend to proceed irrespective of local views and village socioeconomic conditions (Resosudarmo 2005; Cramb and McCarthy, 2016; Robinson and McCarthy, 2016; Gatto et al 2015; Pramudya et al 2017). Although these power dynamics may be concerning from the perspective of indigenous land rights and a host of other reasons, they are helpful for identification since villages have presumably limited ability to affect factory placement. Since districts typically determine whether new factory establishments proceed in their jurisdiction, my counterfactual comparisons are always against unexposed villages nearby, with the same local institutions, politics, and policy settings.

Factories are the processing core for thousands of hectares of surrounding farmland. The primary consideration for investors when placing a new factory is sufficient suitable cultivation area. The optimal factory location in a relatively undifferentiated rural landscape is clearly where yields are highest, maximizing feed quality while minimizing the area needed to operate efficiently (author communications; Corley and Tinker, 2016). Oil palm grows best in the humid low-lying tropics: flat areas with mineral soil, abundant rainfall, and sufficient sunlight. Tree crop estates and transport logistics are also more costly to build and manage in more hilly areas. The agronomic literature and extensive discussions with firms and farmers emphasize the primacy of growing conditions (Corley and Tinker, 2015; Naylor et al, 2019), which firms typically assess based on local rainfall and humidity, overall terrain, and mineral versus peat soil (author communications). As I show shortly, factory placement very closely follows these plausibly exogenous agro-climatic growing conditions, which have been widely used as a source of exogenous identifying variation in economics (Nunn and Qian, 2011; Bustos, Caprettini, and Ponticelli, 2016; Costinot, Donaldson, and Smith, 2016; Gollin, Hansen, and Wingender, 2020).

3 Empirical strategy

Immediate processing implies that local adoption and direct impacts decay over distance. Villages near but not near enough should make a reasonable comparison group since they share many local institutional, economic, and geographic characteristics. Figure 3 uses the PT Tunas Sawa Erma factory in Papua to illustrate the intuition of the spatial difference-type approach.[6] In Panel A, there is just the factory (green dot), village boundaries (light gray lines), and district boundaries (thick black line). Taking the centroids of all the villages, we have the points in Panel B. Every village is assigned an indicator for whether it fits within one of many distance bands, 20 km each in this illustrative example. The red line demarcates 100 km from a factory, beyond which villages are discarded from the estimation sample. Identification is based on spatial heterogeneity and the main identifying assumption is conditional independence: conditional on

[6] See Druckenmiller and Hsiang (2019) for a related discussion.

a comprehensive set of village-specific controls capturing relative within-catchment suitability, this factory could have been placed in any of these distance bands.

3.1 Estimating equation

Factories are related to village outcomes with the equation:

$$y_v = \alpha + \sum_{i=0-5}^{75-80} \gamma dfact_v^i + \sum_{j=1}^{n} \theta fact_v^j + \beta X_v + \epsilon_v \tag{1}$$

y_v is a cross-section of village outcomes measured in censuses, surveys, and other contemporary datasets, detailed further at Appendix A.[7] $dfact_v^i$ are treatment indicators equal to one if a village is 0–5 km, 5–10 km, ... , 75–80 km to the nearest palm oil processor. 80–100 km is the excluded bin, the maximum distance included in the sample to ensure geographic comparability. The universe of palm oil processors is identified through the full 2016 Economic Census, itself an important contribution as other palm oil mill datasets have historically tended to be incomplete. Every processor is exactly matched to the centroid of the village where it is domiciled. The estimation sample excludes villages in cities and on Java, where little oil palm is grown but offices (with the same industry codes as factories) and downstream processors are often located. Nearest factory fixed effects $fact_v^j$ restrict comparisons to villages near the same factory, allowing for level differences and capturing any unobservables across localities.[8]

X_v is a vector of controls capturing the relative attractiveness of each village for a factory. Village elevation, slope, historical average annual precipitation, water flow accumulation (measured as the average number of cells uphill), and the distance to the nearest river account for terrain and relative growing conditions. The distance to the nearest major road in 2000 and nighttime luminosity in 1993 proxy preexisting utilities, industry, and economic development.

[7] Outcomes generally are all measured as close as possible to 2015–16.

[8] Results are similar without these three sample restrictions (i.e., removing cities, village on Java, and village over 100 km from a palm factory; see Figures A11–A13), without nearest factory fixed effects, and with district or subdistrict fixed effects (see Figures A14–A16 and Table 4).

Since roads and utilities are more expensive to build in more rugged terrain further from existing networks, these two variables also proxy the relative cost of new supply chain infrastructure. Distance to the nearest district capital city, export port, and village area further account for administrative and economic remoteness and the relative cost and availability of land, and an urban dummy allows a different intercept for preexisting towns. All are computed before the expansion, and results are similar without conditioning on covariates (see Figures A17–A19). ϵ_v is a robust error term.[9]

Figure 4 maps processor locations and Table 1 provides summary statistics. Restricting the sample to only villages within 100 km of a processor, off Java, and outside cities, the average distance to a factory is 32 km. Twenty percent of villages are within 10 km of a factory and the average village has 2.4 processors within 25 km and 8.3 processors within 50 km. Processing is particularly concentrated in the provinces of Riau and North Sumatra, where the crop was first introduced by the Dutch in the early 20th century.[10]

3.2 Identification

Equation 1 allows us to see how the average value of an outcome varies with proximity to palm factories. Such "between" variation provides a natural long-run interpretation well-suited to studying urbanization and structural change.[11] The identification assumption is conditional independence: that geographic variables related to growing conditions effectively capture factory placement within each catchment, and that any other characteristics tend to be uncorrelated with factory placement and outcomes. To assess the credibility of this assumption, Table 2 conducts

[9]When the observation is at the sub-village level, for example at the level of the individual or household, standard errors are clustered at the village level. Results are similar if clustering at the kecamaten level (the level above the village but below the district, to allow for some spatial correlation within subdistricts) or using Conley errors with spatial decay, so these additional estimates are omitted for brevity.

[10]Additional descriptive figures are provided in the online appendix. Figure A25 provides a heat map of the distance from every village to its nearest processor, Figure A26 maps village palm oil acreage across Indonesia, and Figures A27–29 zoom in on each region, mapping village palm acreage and processor locations.

[11]I stress that treatment effects here are relative differences between similar villages close by. It could be the case that both treatment and control villages are affected by aggregate impacts (e.g., at the district or supply shed level), which would likely lead to underestimating the true effects.

a simple "balance test" comparing villages within 10 km of a palm processor to those 10–100 km away. Panel A contains variables closely related to palm oil growing conditions, Panel B contains remaining geographic variables, and the final column reports the raw (c.f., conditional) normalized differences. The standard rule of thumb for imbalance is a normalized difference greater than 0.25 (Imbens, 2015; Imbens and Rubin, 2015).

The balance test reveals two important patterns. First, suitability-related variables are imbalanced near and farther from factories. These differences are observed for agro-climatically attainable palm oil yields calculated from the Food and Agricultural Organization's Global Agro-Ecological Zones (FAO-GAEZ) data and more finely measured geographic characteristics like altitude, slope, and rainfall, suggesting that factory placement tends to follow growing conditions. In contrast, there are no major differences for geographic variables unrelated to palm growing conditions (e.g., agro-climatically attainable yields for other crops, river, port, and city proximity, travel cost to the nearest city and administrative capital, and village area).[12]

The key identification concern is factories being built in areas already growing, with better preexisting infrastructure, skilled workers, and so forth. However, expansion into largely subsistence regions (i.e., without prior industrial activity) makes this less likely. It is also unclear whether we should a priori expect positive selection. Firms may prefer to target poorer, less developed areas with lower land and labor costs (and possibly more interested or questionable local leaders). To test for purposive placement more formally, I conduct two complementary falsification tests using pre-period outcomes observed in the 1993 Village Census. The first uses my main factory dataset but imperfectly identifies recently treated villages because I restrict the sample to villages in districts not producing palm oil in 2000. The second precisely identifies a subset of recently treated villages (i.e., without relying on aggregate district production), but does not use my main dataset. Specifically, for the much smaller subsample of factories observed in the public Global Forest Watch database, I manually inspect historical LandSat satellite imagery

[12]The exception is coconut suitability: also a palm tree, this is arguably not a major concern. Analogous balance tests estimating Equation 1 using each geographic characteristic as the dependent variable are at Appendix Figures A1 and A2.

for each factory to identify construction dates and create a new auxiliary dataset of 342 factories definitively built after 2000.[13] Table 3 compares the normalized differences for a host of pre-period outcomes using only villages near and farther from new factories identified from the satellite imagery. Raw differences are less than 0.25. Similar results using my main non-linear distance band approach, which adjust for factory fixed effects and initial conditions controls, are plotted in Figures A3–A5, where these small differences dissipate further.

Another identification concern is that there may still be unobserved omitted variables jointly affecting factory placement and outcomes. This threat cannot be ruled out in any non-randomised study, but I provide two types of additional evidence, following the main results, to give the reader additional confidence. First, I introduce a host of additional controls to try and explain away the main results. Second, given the role agricultural suitability in determining placement, I present analogous results from multiple suitability-based IV strategies.

3.3 Proximate adoption and poverty

Using villages slightly farther away from factories as a control group assumes firstly that oil palm adoption and other direct impacts are concentrated around factories and villages slightly farther away are unaffected, or at least much less affected—an implicit "first stage". Figure 5 presents the estimated coefficients at each distance band using village oil palm acreage and estimated poverty (i.e., via small-area estimation techniques) as dependent variables. Each subfigure overlays estimates from (a) the baseline specification, (b) the baseline specification and a host of geographic controls from the 2014 Village Census, and (c) the baseline specification and a full polynomial in latitude and longitude, plus the additional controls in (b), to help purge any remaining geographically-distributed unobservables.[14] The first sub-figure shows how adoption

[13] I use Global Forest Watch rather than the 2016 Economic Census data here because their processor locations are exactly geocoded (c.f., assigned village centroids), making this manual inspection exercise orders of magnitude faster, as villages can be quite large relative to factory size in remote parts of Indonesia. Note also that this is a highly imperfect exercise: many establishment dates could not be identified due to cloud cover or poor image quality, especially before 2004.

[14] The 2014 Village Census geographic controls are the travel time to the nearest city (accounting for road quality), travel cost to the nearest city, and indicators for river, coast, plains, valleys, forest proximity (i.e., in, near, and outside),

decays up to 30 km, when the estimated coefficient becomes statistically indistinguishable from zero. Within 10km of a factory, villages have on average 100 more hectares planted—50 households at the median plot size of 2 hectares.[15] The second sub-figure finds poverty five percent lower near factories. To the extent that (a) causal chains start with local adoption, and (b) other development outcomes are reasonably correlated with poverty, the overlapping confidence intervals, with this rich vector of controls, suggest the spatial patterns I proceed to document are likely to be due to the factories. I stress that this interpretation does not preclude villages within 30–100 km also being affected in aggregate terms, as these are relative comparisons.

4 Main results

4.1 Local labor markets

Figure 6 characterizes the structure of the local economy around factories using the 2010 Population Census.[16] Effects on everyone are in green and women in grey. Short dashes indicate 95 per cent confidence intervals. The first five panels examine employment across sectors, revealing an 8 percent decline in agricultural employment close to factories. Smaller declines are statistically significant at the 5 percent level as far as 20 km away from the factory. Estate crops employment increases by around 4 percent within 5 km. The decline in agricultural employment is reallocated approximately evenly across industry and services, with a four percent increase in each. Spatial decays are similar across sectors and three quarters of the new industrial employment is in manufacturing. The point estimates suggest that men and women are leaving agriculture, but men are more likely to go to industry and women to services. There are minor differences for women, although all confidence intervals overlap except industry.

and forest function, e.g., production or conservation).

[15]Note that the common industry rule of thumb for a supply shed is 50km from the mill, which may be overstating the extent to which farmers are willing to travel and transport their fruit.

[16]The 2010 Population Census is more precise than newer datasets potentially allowing early onset effects of post-2010 factories to be detected. However, most factories were there before 2010 and pooled SUSENAS or the 2014 Village Census both yield similar results (Figures A6 and A7).

The remaining remaining panels of Figure 6 look at employment status. Within 5 km of a factory, people are over 15 per cent more likely to be employed. Most of the effect is driven by permanent employment. The increase in employment is mirrored by a decline in self-employment and housework, and the decline in housework is almost twice the size of the decline in self-employment. Note that agricultural and informal employment is typically reported as self-employment. The increase in permanent employment and decrease in self-employment are driven by men, with effects on women statistically indistinguishable from zero. Declines in housework are statistically indistinguishable across genders but the point estimates are larger for women. Social norms could see men say they are self-employed or entrepreneurs while women report doing housework, particularly if enumeration is done for the whole household. In practice, both may be active on the farm, family business, and home, suggesting a cautious interpretation. Effects on work status are more precisely estimated and larger than those on primary sector of employment and expenditures, perhaps reflecting the fact that people outside formal employment are more likely to have multiple income sources. Overall, the shift out of self-employment and agriculture we observe here is consistent with the spatial division of labor usually observed around nascent urban centers (Fafchamps, 2012).

One possible concern with new factories and plantation-related infrastructure altering the structure and nature of local employment is that the new jobs may actually decrease household welfare. An extreme but not unrealistic example is that agricultural households may become landless, left with little choice but to work for a monopsonistic employer for wages lower than what they could previously earn on their own farms. Such a story of immersing growth is often told about large commercial farms and plantation economies (Easterly, 2007; Byerlee et al., 2009). I test this conjecture by examining consumption (more specifically, expenditure) patterns around factories. Household expenditures are measured in SUSENAS, a repeat cross-sectional household survey conducted at least annually. I follow Dell and Olken (2019) and pool data over consecutive years with village identifiers to improve coverage. Household size, an urban-rural dummy variable, and survey-year fixed effects are included as additional controls throughout, and standard errors are clustered by village. Per capita household expenditures, for all households, are on average ten percent higher within 10 km of a palm processor (see Figure A8). Impacts decay linearly up to 25 km away. Estimating impacts separately by the primary source of income reveals similar point estimates across sectors, but the aggregate effects appear driven by agriculture (the main source of employment). Services impacts are more concentrated, decaying only up to 15 km. There is no clear pattern for households employed in industry, consistent with skilled labor being more mobile and wage equalization across locations.

4.2 Population and firms

Living standards tend to improve with density and the non-agricultural sector is often more productive (Gollin et al., 2014; Gollin, Kirchberger, and Lagakos, 2021). Reallocation and higher incomes should be mirrored by changes in population and firms. Figure 7 estimates impacts on village population and whether a village is home to a formal registered firm, measured in the 2011 Village Census and the 2016 Economic Census. I overlay results including a complete polynomial in latitude and longitude (i.e., the most spatially demanding specification). The two specifications yield remarkably similar patterns. Magnitudes are not trivial. The average village 5-10 km from a palm oil factory has 500 more people and is fifteen percent more likely to be home to a formal firm. On the intensive margin, this equates to approximately an additional firm per village. Effects are roughly twice as large within 5 km.

The shift out of agriculture into formal employment may be complicated by population growth, as skilled in-migration could fully explain changes in employment. To explore this possibility, I examine impacts separately for "locals" and recent migrants in Figure A9. Recent migrants are defined as people who lived in a different district in 2005, as reported in the 2010 Population Census. Effects on employment status are not statistically distinguishable by migration status, although the declines in self-employment and housework appear driven by non-migrants. The decline in agricultural employment is also concentrated amongst locals, although estate crop workers are more likely to be migrants. Results are qualitatively similar using inter-province migration or lifetime migration status (i.e., place of birth, rather than place of residence five years ago). Hence, the changes in employment appear unlikely to be driven by selective in-migration.

Results presented up to now show a strong association between processor proximity and a host of outcomes synonymous with nascent urban centers. However, these patterns can only be interpreted as the causal effects of the proliferation of palm oil factories under conditional unconfoundedness. Although I have shown that factory placement closely follows suitability and not other geographic characteristics or pre-period outcomes, one might still be concerned that the patterns could be explained by some unobserved variable correlated with factory locations and outcomes. Here, I briefly introduce a host of additional controls and present analogous results from three different IV approaches to ease such concerns.

4.3 Robustness—additional controls

To present estimates with additional covariates more compactly, I switch to a dichotomous treatment indicator equal to one if villages are less than 10 km from a factory and focus only on the primary outcomes of population, firms (extensive margin), and whether a village's main source of income is agriculture. Factory fixed effects and geographic controls remain, so observationally similar villages 10–100 km away are the comparison group.[17] Column 1 of Table 4 presents the ordinary least squares (OLS) estimate without conditioning on any controls (i.e., including only nearest factory fixed effects). Column 2 is the binary treatment analogue of Figure 7. Column 3 adds additional geographic variables reported in the 2014 Village Census (listed at Figure 5) and Column 4 adds a polynomial in latitude and longitude. Column 5 adds each village's agro-climatically attainable palm oil yield calculated from the FAO-GAEZ data.[18] Column 6 alternates the nearest-factory fixed effects for district fixed effects. Column 7

[17] Although villages 10–30 km away could be directly affected, results are similar if a "donut hole" type approach is used to remove these villages. This additional check, omitted here for brevity, suggests spatial sorting within supply sheds is unlikely to be a major concern.

[18] Although GAEZ potential yields lacks the precision of my baseline suitability variables, the model-based yield response function adds additional crop-specific information.

adds subdistrict fixed effects to exploit only within-subdistrict differences.[19] Estimates are not statistically distinguishable across these seven specifications.

4.4 Robustness—instrumental variables

An appropriate IV will predict factory exposure but not affect outcomes through any other channel other than factory proximity. My IVs leverage the fact that palm factories locate in the most suitable areas. The first follows Duflo and Pande (2007), Dinkelman (2011), and Lipscomb et al (2013), using a simple probit model to predict the probability a village will be within 10 km of a factory. Predictions are made purely based on exogenous determinants of profitability and predicted probabilities instrument actual exposure. The granularity of the geospatial data is important, offering more variation within localities than the standard FAO-GAEZ suitability data. However, FAO-GAEZ data has the important benefit of incorporating crop-specific yield functions from agronomic models without using any information on observed productivity and other human activity on the ground (Fischer, van Nelthuizen, Shah, and Nachtergaele, 2002). My second instrument is thus village agro-climatically attainable palm oil yield calculated from the FAO-GAEZ data. The key concern with my first two instruments is that palm-suitable areas might be also suitable for other types of agriculture, so third instrument, following Lowes and Montero (2021), takes the normalized difference between potential yields for palm oil and other key cash crops—specifically coffee, cocoa, and teas—to directly address this concern.

Table 5 present the results from the three IV approaches. IV estimates are larger than the corresponding OLS estimates in Table 4, reflecting a local average treatment effect larger than the average treatment effect (i.e., effects will likely be stronger if you follow an optimal placement rule), a downwards OLS bias (e.g., due to firms targeting areas with worse unobservable village characteristics, like corruption), or some combination of both.

[19] Subdistricts, known in Indonesian as kecamaten, are the administrative level above a village but below a district (or kabupaten, or regency), usually much smaller than a factory catchment.

The crucial IV identification assumption is that palm suitability is conditionally independent, not affecting outcomes other than through factory exposure. The concern is that more suitable areas were going to urbanize anyway, for some reason other than agricultural processing.[20] I conduct an intuitive falsification test to explore this possibility. The first panel of Table 6 presents the reduced form impacts on population, firms, and agriculture. The second panel presents the same for a restricted sample discarding villages within 25 km of a factory, where the first stage is effectively zero. There is no evidence of any statistically significant effect of palm suitability or relative suitability in unexposed villages, suggesting that palm cultivation and processing is the principal channel through which palm suitability affects outcomes.

5 Potential explanations

5.1 Economic linkages

The main results suggest that the location of palm oil production affects other economic activity through agglomeration economies and other forms of economic spillovers. I now examine three potential explanations for these patterns. Armed with the knowledge that the average village 5-10 km from a factory has one additional firm present, a natural first question to ask concerns economic linkages. Do these neighboring firms provide services to palm oil processors and farmers, or use palm-related products as inputs?

I use data on the universe of formal firms in Indonesia in 2016 to examine the relationship between factory proximity and the presence of different types of firms. Sectors are identified by industry codes (KBLI 2015) in the 2016 Economic Census. Linked industries are identified in the 2010 Input-Output Table. Since codes change over time and there is not yet a newer input-output table or official concordance, I translate the sector descriptions from Indonesian to English then match sector codes one-by-one based on their descriptions.[21]

[20] The main distance band estimates address this concern by adjusting all estimates for suitability.

[21] Appendix B provides details. Note that I was unable to link 9 out of 51 of the input sectors due to translation issues, so the coefficients for input linkages and relative differences are best interpreted as lower bounds. All output

Figure 8 uses the primary distance band specification to relate palm oil factories to other economic activities. The top left panel finds villages near factories slightly more likely to be home to firms providing inputs to the palm oil sector. Effects are strongest within 5 km and remain statistically significant at the 5 per cent level up to 20 km away. I find no evidence of forward (i.e., output) linkages. Point estimates are quantitatively small and indiscernible from zero at conventional levels of statistical significance. Impacts on linked industries overall, in the top right panel, are thus driven by backward linkages. The next panel (first column, second row) examines whether villages near factories have more firms in sectors without production linkages. Villages near palm factories are around ten times more likely to have a firm in a non-linked industry than one in a linked industry. Figure 9 disaggregates these patterns further by estimating impacts by sector. Again, I replace the distance bands with a 10 km dichotomous treatment indicator. Villages near factories have more retail, maintenance, finance, insurance, transport, construction, and other service sector firms.

The key limitation of assessing rural economic activity using the Economic Census is that most micro and small enterprises in rural areas are not registered and not counted. These smaller organizations are more common, particularly at early stages of economic development (Hsieh and Olken, 2014; Rothenberg, Gaduh, Burger, Chazali, Tjandraningsih, Radijun, Sutera, and Weilant, 2016). The 2014 Village Census allows me to probe local economic activity further, as village heads are asked how many businesses and organizations are in their village, regardless of formal status. The remaining five panels of Figure 8 present these estimates. I find no evidence of more micro-industry or small-scale processing, suggesting that industrial employment is most likely coming through formal firms, including mechanically through the factories. Consistent with Figure 9, a village from 5-10 km from a factory has, on average, ten more small service businesses. Villages near factories are also more likely to have cooperatives (usually used to organize agricultural activities) or a bank, and people living in villages moderately close are more likely to have received credit in the last year. Collectively, these findings suggest that consumption

sectors, however, were successfully linked.

linkages, rather than production linkages, are more likely to underpin the agglomerations.

5.2 Infrastructure and other public goods

As palm expansion has mostly been into remote areas with little commercial infrastructure, firms and governments seeking to develop a local palm industry may need to invest in infrastructure to transport, process, and market the crop. Such infrastructure could reinforce agglomeration and make rural public good provision easier, offering new revenue opportunities while reducing the cost of service provision. Local industrial activity could also raise the returns on public investments, like education and infrastructure. I find that villages near factories do indeed have greater fiscal capacity to provide such public goods (Figure A10). Specifically, village near factories have over fifty percent higher own source revenues, receive twenty percent less in Alokasi Dana Desa (ADD) intergovernmental grants (which are based partially on need), and have around ten percent higher expenditures. Villages near factories are also more likely to report having land (7 percent more likely), building (12 percent), and other assets (5 percent).

Figure 10 turns to public goods. Across outcomes, impacts vary in magnitude, statistical significance, and the distance at which they are felt. The top left figure shows that villages living within 10km of a palm oil factory tend to have 10 less families living without access to any electricity, despite having larger populations. Rural electrification in Indonesia is generally high, and population in 2011 is included as a control variable. The second and third figures (top row) find villages within 20km of a factory more likely to have the main road fitted with a street light, and the main road is less likely to be dirt but rather upgraded to gravel or asphalt. The first figure in the second row finds villages within 5 km of factories slightly more likely to have public transport. The middle and right figures (middle row) estimate impacts on the total number of physical marketplaces, crucial centers of rural exchange. Up to 20 km from a factory, villages are five percent more likely to have a market. Just 16 percent of villages had markets in 2014, according to the Village Census.

The bottom row of Figure 10 looks at infrastructure more related to human development than agricultural supply chains. I find large and precisely estimated impacts on the number of education, health, and worship facilities.[22] Impacts are statistically significant at the 5% level as far as 25 km away. Although new supply chains could generate mechanical improvements in economic infrastructure, impacts on health and education facilities suggest public goods spillovers beyond the supply chain. To gauge whether village fiscal windfalls can help explain the increased public good provision, I overlay in red estimates including local government revenue and expenditure as additional covariates. The point estimate relating to having a dirt road is the only one with noticeable movement, although the 95% confidence intervals still overlap. Despite fiscal variables being “bad controls”, these statistically indistinguishable results hint that public goods may be privately provided or the result of some form of intergovernmental transfer.

5.3 Scale economies in production

Improved infrastructure opens up the possibility of falling trade costs (broadly defined) and better market integration as additional channels at work. Up-front infrastructure investments made by one firm and subsequent smallholder adoption could make it attractive for more processors to set up operations nearby. Since most smallholders sell through intermediaries and aggregators, thicker farm-to-factory markets could also encourage smallholder entry. Economies of scale in palm production and processing could thus reinforce agglomerations, allowing potential gains to ratchet up over time.[23]

I test for scale economies with a simple OLS exercise. To proxy local farm-to-factory market density, I calculate in GIS the number of processors each village could plausibly sell palm fruit to based on 25 and 50-km buffers. 50 km is the standard industry definition of a supply shed,

[22]I look at places of worship because mosques and churches are often the first thing communities invest in (a helpful leading indicator of community financial health), and because they often serve as important centers of social activity, social services (e.g., child care), and risk-sharing.

[23]Ideally we would also test for internal economies of scale by comparing small and large processing facilities. However, Indonesian palm oil processors are relatively homogenous in size and capacity, mostly coded as large firms in the Economic Census.

but 25 more closely resembles the smallholder adoption patterns in Figure 5. I augment my main specification with this single continuous treatment indicator (c.f., the distance bands), exploiting only heterogeneity in farmer-to-factory market competition across similarly suitable villages within each factory catchment.

Table 7 presents the results. Panel A considers whether a village's primary source of income is agriculture and total village population. Column 1 shows that an additional buyer within 25 km corresponds an 0.3 percent decrease in the probability that most village income comes from agriculture. Column 2 adds the 10 km proximity treatment and an interaction term with potential buyers. The direct effect of additional buyers becomes statistically significant and the magnitude small. The interaction term is statistically significant at the 1 percent level, suggesting that improved local market integration may be reinforcing the shift out of agriculture, but only very close to the factories. Put differently, the proximity effects appear largest for the subset of communities near processors which also have other processors in marketing distance. Columns 3 and 4 show similar results using a 50 km buffer. Coefficients on potential buyers are smaller, as expected. Columns 5–8 of Panel A examine population. For the 25 km buffers, both direct effects and the interaction term are large and statistically significant. For the 50 km buffers, the direct effect of market density is rendered statistically insignificant. [24]

Panel B of Table 7 turns to firms. I present results for the number of firms per village and whether there is a firm in the village, painting a slightly more nuanced picture than either outcome alone.[25] On the intensive margin, being within 10km of factory, having more factories within 25 km, and the interaction of these variables are all statistically and economically significant. For the larger buffer, effects are weaker, proximity dominates, and the interaction term is statistically insignificant. A slightly different picture emergences on the extensive margin. While more potential buyers at either margin corresponds to an increased probability of having

[24]Taking the natural log of population to reduce the influence of outliers gives similar results and similar patterns are also observed using the inverse distance to the nearest processor or a non-linear distance band specification. These estimates are omitted for brevity.

[25]Note that all other firm estimates are similar in direction and precision whether using the count or binary outcome.

another firm (excluding palm firms) nearby, the proximity treatment renders the coefficient on market density much smaller and statistically indistinguishable from zero. [26] I cautiously interpret these estimates as suggesting that farm-to-factory market density is more helpful for growing economic activity in areas that already have good market access and firms present, rather than encouraging first-movers.

6 Conclusion

This paper used the recent proliferation of Indonesian palm oil factories across its underdeveloped hinterland to estimate spillovers from agricultural processing. The key finding is that new export-oriented agricultural processing factories are reshaping Indonesia's rural economic geography, generating economic spillovers well beyond those accruing directly to farmers. Living in a village near a palm oil factory corresponds to a greater likelihood of non-agricultural employment and higher household incomes. Villages near factories also have more people, large firms, and other economic and social organizations. These effects appear driven by income effects and consumption linkages, rather than linked industries, as in Dell and Olken (2019). Villages near factories also tend to generate significantly more revenue, hold more assets, and have better roads, increased electrification, and more marketplaces, schools, and health facilities. Consistent with improved local infrastructure reducing trade costs and improving local market integration, the agglomerations appear to be strongest for villages with thicker farm-to-factory markets.

By focusing on reduced-form comparisons across villages, my analysis precludes any conclusions on whether Indonesia's dramatic palm oil expansion supports or hinders aggregate structural transformation. In particular, I cannot infer whether reinforcing Indonesia's comparative advantage in agriculture and relatively low-skilled manufacturing—which might

[26] One possible concern is that intensive margin effects are driven by high-count, high-leverage observations treated the same with extensive margin binary outcomes. Similar effects taking logs or using an inverse hyperbolic sine transformation suggest that this is unlikely to be the case here.

undertake less research, development, and innovation—will lead to lower long-run growth (Matsuyama, 1992; Bustos, Vincenzi, Monras, and Ponticelli, 2022). Rather, this paper found highly localized patterns of agglomeration, urbanization, and economic development around factories in otherwise remote, rural areas. Modern agricultural supply chain investments appear to be planting the seeds of broader economic development deep in the Indonesian countryside, offering several exciting new research directions. For example, future work might structurally estimate the general equilibrium impacts, explore linkages between major cities and rural areas in more detail, or delve deeper into other long-run implications.

The findings in this paper inform topical policy debates on global trade, food systems, and economic development. While there remains significant work to be done to improve the overall environmental sustainability of our global food system, the Indonesian case highlights the potential of export-oriented agricultural manufacturing as an avenue for developing countries to better integrate into global trade networks, attract investment, and improve livelihoods in rural regions with otherwise limited economic opportunities.

References

Allen, T., and Arkolakis, C. (2014), "Trade and the Topography of the Spatial Economy", Quarterly Journal of Economics, 129: 3, pp. 1085–1140.

Barrett, C.B., Reardon, T., Swinnen, J., and Zilberman, . (2022) "Agri-food Value Chain Revolutions in Low- and Middle- Income Countries", Journal of Economic Literature, 60:4, pp. 1316–77.

Bellemare, M., and Bloem, J. (2018), "Contract farming: a review," World Development, 112, December 2018, pp. 259–271.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016), "Agricultural Productivity and Structural Transformation: Evidence from Brazil," American Economic Review, 106:1, pp. 130–1365.

Bustos, P., Castro-Vincenzi, J., Monras, J., and Ponticelli, J. (2022) "Industrialisation without innovation', working paper.

Byerlee, D., de Janvry, A., Sadoulet, E. (2009), "Agriculture for Development: Toward a New Paradigm", Annual Review of Resource Economics", 1: 15–31.

Byerlee, D., Falcon, W., and Naylor, R. (2016), "The Tropical Oil Crops' Revolution: Food, Feed, Fuel, and Forests", Oxford University Press, Oxford.

Chenery and Syrquin (1975), "Patterns of Development, 1955–1970," Oxford University Press, London.

Clark, C. (1940), "The Conditions of Economic Progress", McMillan and Co., London.

Corley, R., and Tinker, R. (2015), "The Oil Palm", 5th Edition, Blackwell, Oxford.

Costinot, A., Donaldson, D., Smith, C. (2016), "Evolving Comparative Advantage and the Impact of Climate Change on Agricultural Markets: Evidence form 1.7 Million Fields around the World," Journal of Political Economy, 24:1, pp. 205–248.

Cramb, R. and McCarthy, J. (2016), "The Oil Palm Complex: Smallholders, Agribusiness and the State in Indonesia and Malaysia", National University of Singapore Press.

Dell, M., and Olken, B. (2019), "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java", Review of Economic Studies, March 2019.

Dercon, S. and Gollin, D. (2014), "Agriculture in African Development: Theories and Strategies", Annual Review of Resource Economics, 6, pp. 471–492

Dinkelman, T. (2011) "The Effects of Rural Electrification on Employment: New Evidence from South Africa", American Economic Review, 101:7, pp. 3078–3108.

Directorate General of Estate Crops (2017), "Tree Crop Estate Statistics of Indonesia, Palm Oil, 2015–2017", Department of Agriculture, December 2017, Jakarta.

Donaldson, D. (2015), "The Gains from Market Integration", Annual Reviews in Economics, 7: 617-47.

Donaldson, D., (2018), “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure, American Economic Review, 108:4-5, pp. 899–934.

Dragusanu, R., Giovannucci, D., and Nunn, N. (2014), “The Economics of Fair Trade,” Journal of Economic Perspectives, 28: 3, pp. 217–236.

Druckenmiller, H. and Hsiang, S. (2019) “Accounting for unobservable heterogeneity in cross section using spatial first differences”, NBER Working Paper 25177, June 2019.

Duflo, E., and Pande, R. (2007) “Dams”, Quarterly Journal of Economics, 2:1, pp. 601–646.

Easterly, W. (2007), “Inequality does cause underdevelopment: Insights from a new instrument,” Journal of Development Economics, 85, pp. 755–776.

Edwards, R. B (2019), “Export agriculture and rural poverty: evidence from Indonesian palm oil”, working paper.

Engerman, S., and Sokoloff, K. (2002) “Factor endowments, inequality, and paths of development among new world economies,” NBER Working Paper Series 9259.

Fafchamps (2012), “Development, Agglomeration, and the Organization of Work”, Regional Science and Urban Economics, 42, pp. 459-472.

Fafchamps, M. Koelle, M. and Shilpi, F. (2017) “Gold Mining and Proto-Urbanization: Recent Evidence from Ghana”, Journal of Economic Geography, 17:5, pp. 795–1008.

Farina, E., and Reardon, T. (2000), “Agrifood Grades and Standards in the Extended Mercosur: Their Role in the Changing Agrifood System,” American Journal of Agricultural Economics, 82:5, pp. 1170–1176.

Fischer, G., van Nelthuizen, H., Shah, M., and Nachtergaele, F. (2002), “Global Agro-ecological Assessment for Agriculture in the 21st Century: Methodology and Results,” Food and Agriculture Organisation, Rome.

Fitriani, F., Hofman, B., and Kaiser, K. (2005), “Unity in diversity? The creation of new local governments in decentralising Indonesia", Bulletin of Indonesian Economic Studies, 74: 1, pp. 70–93.

Fujita, M., Krogman, P., and Venables, A.J. (1998), “The Spatial Economy: Cities, Regions, and International Trade,” MIT Press, Cambridge MA.

Gaskell, J. C. (2015), “The Role of Markets, Technology, and Policy in Generating Palm-Oil Demand in Indonesia", Bulletin of Indonesian Economic Studies, 51:1, pp. 29–45.

Gatto, M., et al (2015), “ Oil palm boom and land-use dynamics in Indonesia: The role of policies and socioeconomic factors”, Land Use Policy, 46, pp. 292–303.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010) “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” Journal of Political Economy, 118:3, pp. 536–498.

Gollin, D., Kirchberger, M., and Lagakos, D. (2021), “Do urban wage premia reflect lower amenities? Evidence from Africa”, Journal of Urban Economics, 121.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014), "The agricultural productivity gap," Quarterly Journal of Economics, 129:2, pp. 939–993.

Gollin, D. (2010), "Agricultural Productivity and Economic Growth", in Handbook of Agricultural Economics, pp 3825–3866.

Gollin, D., Hansen, C.W., and Wingender, A.M. (2020) "Two Blades of Grass: The Impact of the Green Revolution", Journal of Political Economy, 129:8.

Hayami, T. (2010), "Plantations Agriculture", in Handbook of Agricultural Economics, pp. 3305–3322.

Hirschman, A.O. (1960), "The Strategy of Economic Development," Yale University Press.

Hsiao, A. (2024) "Coordination and Commitment in International Climate Action: Evidence from Palm Oil", working paper.

Hsieh, C., and Olken, B. A. (2014), "The missing "Missing Middle"", Journal of Economic Perspectives, 28:3, pp. 89–108.

Imbens, G.W. (2015) "Matching Methods in Practice: Three Examples", Journal of Human Resources, 50:2 pp. 373–419.

Imbens, G.W., and Rubin, D.B. (2015) "Causal Inference for Statistics, Social, and Biomedical Sciences", Cambridge University Press.

Johnston, B.F., and Mellor, J.W. (1961), "The Role of Agriculture in Economic Development", American Economic Review, 51:4, pp. 566–593.

Krishna, V. et al. (2017) "Differential livelihood impacts of oil palm expansion in Indonesia", Agricultural Economics, 48, 639–653.

Lipscomb, M., Mobarak, A.M., and Barham, T. (2013) "The development effects of electrification: evidence from the topographic placement of hydropower plants in Brazil", American Economic Journal: Applied Economics, 5:2, pp. 200–23.

Lowes, S., and Montero, E. (2021) "The Legacy of Colonial Medicine in Central Africa", American Economic Review, 111:3, pp. 1284–1314.

Maertens, M. and Swinnen, J.F.M. (2009), "Trade, Standards, and Poverty: Evidence from Senegal," World Development, 37:1, pp. 161–178.

Marden, S. (2018) "The Agricultural Roots of Industrial Development: Rural Savings and Industrialisation in Reform Era China," working paper.

Matsuyama, K. (1992), "Agricultural Productivity, Comparative Advantage, and Economic Growth," Journal of Economic Theory, 58: 2, pp. 317–34.

Moretti, E. (2010) "Local labor markets," Handbook of Labor Economics, Volume 4b.

Michaels, G., Rauch, F., and Redding, S. J. (2012), "Urbanization and structural transformation," Quarterly Journal of Economics, 127:2, pp. 535–586.

Myrdal, G. (1957) "Economic Theory and Under Developed Regions", G. Duckworth.

Naylor, R. L., Higgins, M. M., Edwards, R. B., and Falcon, W. P. (2019), “Decentralization and the Environment: Assessing Smallholder Oil Palm Development in Indonesia”, Ambio, January, pp. 1–14.

Nunn, N., and Qian, N. (2011) “The Potato’s Contribution to Population and Urbanization: Evidence from a Historical Experiment”, Quarterly Journal of Economics, 126, pp. 593–650

Pramudya, E. P., Hospes, O., and Termeer, C. (2017), “Governing the palm oil sector through finance: the changing roles of the Indonesian state,” Bulletin of Indonesian Economic Studies, 53:1, pp. 57-82.

Pryor, F. (1982), “The Plantation Economy as an Economic System," Journal of Comparative Economics, 7, pp. 288–317.

Rada, N., Buccola, S.T., and Fuglie, K.O. (2011), “Government policy and agricultural productivity in Indonesia”, American Journal of Agricultural Economics, 93:3, pp. 867–884.

Resosudarmo, B. (2005) “The Politics and Economics of Indonesia’s Natural Resources,” Indonesia Update Series, Research School of Pacific and Asian Studies, The Australian National University, ISEAS Press, Singapore.

Robinson, K., and McCarthy, J. (2016) “Land and development in Indonesia: searching for the people’s sovereignty,” Indonesia Update Series, College of Asia and the Pacific, The Australian National University, ISEAS Press, Singapore.

Rothenberg, A., Gaduh, A., Burger, N., Chazali, C., Tjandraningsih, I., Radijun, R., Sutera, C., and Weilant, S. (2016), “Rethinking Indonesia’s Informal Sector,” World Development, 80, 96–113.

Schultz, T.W. (1964), “Transforming Traditional Agriculture", Yale University Press, New Haven, Connecticut.

Swinnen, J.F.M. (2007), “Global supply chains, standards, and the poor,” CABI Publishing, Oxford, UK.

World Bank (2008), World Development Report: Agriculture for Development, Washington, D.C.

Tables and Figures

Table 1: Descriptive Statistics—Palm Oil Processors

Mean	Mean	SD	Min	Median	Max
Estimation sample					
Distance to nearest factory (km)	32.3	25.7	0	26.03	99.99
Within 10km of a factory (=1)	0.19	0.40	0	0	1
Factories within 25km (n)	2.4	6.9	0	0	82
Factories within 50km (n)	8.3	16.3	0	3	106
N = 30,330 villages					
Full sample					
Distance to nearest factory (km)	93.7	126.4	0	48.5	809.9
Within 10km of a factory (=1)	0.115	0.319	0	0	1
Factories with 25km (n)	1.9	7.2	0	0	82
Factories within 50km (n)	5.6	14.1	0	1	106
N = 81,695 villages					

Notes: The dataset for these summary statistics was prepared by the author in geographic information systems (GIS) using 2016 village boundaries and the 2016 Economic Census. All distance calculations use the Asia South Albers Equal Area projection with minor adjustments. The estimation sample drops villages in cities, on Java, and beyond 100km of a factory, and is the sample used when examining population (note that estimation samples vary across outcomes depending on data availability and match success).

Table 2: Balance on Geographic Characteristics—Raw Unadjusted Differences

	Control (10-100 km)		Treatment (<10 km)		Normalized
Variable	Mean	S.E.	Mean	S.E.	difference
Panel A—Palm Suitability					
Potential palm oil yield/ha	3789.011	[12.170]	4717.279	[7.902]	0.471
Slope	2.923	[0.021]	1.001	[0.021]	0.565
Elevation	283.615	[2.396]	66.304	[1.417]	0.560
On a plain (=1)	0.777	[0.002]	0.939	[0.003]	0.410
Total precipitation in 2015	2094.816	[3.236]	2188.266	[5.967]	0.169
Average long term precipitation	2584.037	[3.262]	2684.298	[5.991]	0.180
Water flow accumulation	459.670	[15.651]	553.749	[30.068]	0.035
Distance to nearest main road in 2000	4921.794	[74.261]	3226.396	[90.666]	0.140
Panel B—Other Geographic Characteristics					
Potential casava yield	5338.555	[12.262]	5422.291	[15.276]	0.042
Potential coconut yield	1689.864	[9.219]	2336.139	[7.133]	0.432
Potential cocoa yield	1430.110	[7.185]	1517.913	[6.259]	0.076
Potential tobacco yield	73.865	[6.195]	132.890	[5.100]	0.059
Potential tea yield	112.639	[6.435]	-1.938	[4.818]	0.111
Potential wetland rice yield	4466.087	[10.139]	4538.414	[10.934]	0.044
Potential coffee yield	1228.021	[6.960]	1265.678	[6.113]	0.034
Potential maize yield	3046.817	[8.310]	3117.034	[7.655]	0.053
On a river (=1)	0.795	[0.002]	0.735	[0.005]	0.148
Distance to nearest river	619.727	[16.062]	365.213	[11.448]	0.099
Village area	3.29e+07	[5.48e+05]	2.59e+07	[8.97e+05]	0.076
Distance to nearest port	11.885	[0.027]	12.796	[0.048]	0.197
Distance to district capital	64.874	[1.964]	56.788	[3.086]	0.025
Cost to district capital	270.314	[14.252]	186.134	[30.383]	0.034
Distance to nearest capital	3.100	[0.026]	2.423	[0.047]	0.154
Cost to nearest capital	392.932	[21.220]	269.118	[43.393]	0.034

Notes: Villages near palm oil factories are identified using 2016 village boundaries and the the Global Forest Watch palm oil mills dataset. All distance calculations use the Asia South Albers Equal Area projection, with minor adjustments. The sample here is the primary estimation sample and drops villages in cities, on Java, and beyond 100km of a factory. Data are described in detail in Appendix A. Complementary balance tests using my main specification, adjusting for factory fixed effects and other covariates, are at Figures A1 and A2.

Table 3: Balance on Pre-Period Outcomes—Raw Unadjusted Differences

	Control (10-100 km)		Treatment (<10 km)		Normalized
Outcome in 1993	Mean	S.E	Mean	S.E	difference
Primary income is agriculture (=1)	0.956	[0.002]	0.928	[0.007]	0.133
Primary income is industry (=1)	0.005	[0.001]	0.009	[0.002]	0.062
Primary income is plantations (=1)	0.167	[0.003]	0.199	[0.010]	0.087
Village population (n)	1458.865	[51.479]	3271.405	[558.664]	0.199
Cooperatives (n)	0.327	[0.013]	0.451	[0.042]	0.087
Road outside village asphalt (=1)	0.382	[0.004]	0.416	[0.013]	0.069
Road outside village soil (=1)	0.227	[0.004]	0.229	[0.011]	0.003
Road inside village asphalt (=1)	0.222	[0.004]	0.194	[0.010]	0.070
Road inside village soil (=1)	0.549	[0.004]	0.521	[0.013]	0.057
Permanent marketplace (=1)	0.076	[0.002]	0.103	[0.008]	0.1
Nonpermanent marketplace (=1)	0.106	[0.003]	0.145	[0.009]	0.124
Households with electricity (%)	0.259	[0.003]	0.305	[0.008]	0.154

Notes: Villages near new palm oil factories are identified using 2016 village boundaries and the the Global Forest Watch palm oil mills dataset. All distance calculations use the Asia South Albers Equal Area projection, with minor adjustments. The sample here is the primary estimation sample drops villages in cities, on Java, and beyond 100km of a factory. Only factories identified visually in satellite data as having been constructed since 2000 are retained in the sample. Outcomes are observed in the 1993 Village Census. Pre-period outcome placebo tests using my main specification, adjusting for factory fixed effects and other covariates, are at Figures A3–A5.

Table 4: OLS estimates—Factory Proximity, Agglomeration, and Agricultural Employment

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Population (n)							
Near factory treatment (=1)	502.163***	447.850***	428.245***	416.085***	419.011***	532.581***	313.284***
	(36.423)	(34.662)	(35.073)	(35.194)	(35.359)	(40.464)	(42.971)
Village observations	32,203	30,296	29,991	29,991	29,759	29,746	29,762
Panel B: Firms present (=1)							
Near factory treatment (=1)	0.145***	0.138***	0.129***	0.127***	0.127***	0.129***	0.119***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)
Village observations	37,022	34,236	33,880	33,880	33,602	31,112	33,595
Panel C: Primary source of village income is agriculture (=1)							
10 km treatment (=1)	-0.033***	-0.025***	-0.023***	-0.023***	-0.023***	-0.021***	-0.016***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Village observations	36,238	34,236	33,880	33,880	33,602	31,112	33,595
Nearest factory FEs	Y	Y	Y	Y	Y	N	N
Baseline geographic controls	N	Y	Y	Y	Y	Y	Y
Additional PODES controls	N	N	Y	Y	Y	Y	Y
Latitude-longitude polynomial	N	N	N	Y	Y	Y	Y
Potential palm oil yield	N	N	N	N	Y	Y	Y
District FEs	N	N	N	N	N	Y	N
Subdistrict FEs	N	N	N	N	N	N	Y

Notes: Sample is a cross-section of all Indonesian villages within 100 km of a palm oil processor, excluding villages in cities and on Java. The treatment variable is set to 1 if a village is within 10 km of a palm factory and 0 otherwise. Changes in sample sizes are due to imperfect village matches across datasets. Baseline geographic controls include elevation, slope, historical precipitation, flow accumulation, distance to river, distance to major road in 2000, nighttime luminosity in 1993, distance to city, village area, and an urban dummy. Additional Village Census (PODES) controls include travel time and travel cost to the nearest city, travel cost to the nearest city; river, coast, plain, and valley dummies; in, near, and outside forest dummies, and conservation and production forest dummies. A polynomial in latitude and longitude refers to the latitude and longitude of village centroids and their squared terms. Potential palm oil yield is calculated from the FAO-GAEZ gridded dataset. Village, subdistrict, and district identifiers are official 2016 definitions in the 2016 BPS shapefile, and the nearest factory to every village is calculated in GIS using the 2016 Economic Census. Population is taken from the 2011 Village Census, firms from the 2016 Economic Census, and primary village income from the 2014 Village Census. Robust standard errors are in parentheses and stars denote statistical significance at the 10 (*), 5(**), and 1(***) percent levels.

Table 5: IV Estimates—Factory Proximity, Agglomeration, and Agricultural Employment

Instrumental variable	Pr (treatment=1)		GAEZ palm suitability		GAEZ relative suitability	
Column	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Population (n)						
Near factory treatment (=1)	1903.676***	1912.715***	836.744***	822.942***	962.602***	958.383***
	(143.695)	(142.926)	(154.404)	(156.095)	(132.876)	(135.331)
Excluded F	827	832	875	827	1,073	1,018
Village observations	30,288	30,288	30,178	30,178	30,178	30,178
Panel B: Firms present (=1)						
Near factory treatment (=1)	0.393***	0.400***	0.195***	0.207***	0.232***	0.248***
	(0.040)	(0.039)	(0.054)	(0.053)	(0.045)	(0.044)
Excluded F	931	951	950	927	1,172	1,145
Village observations	34,226	34,226	34,080	34,080	34,080	34080
Panel C: Primary source of village income is agriculture (=1)						
Near factory treatment (=1)	-0.114***	-0.109***	-0.041*	-0.034	-0.042**	-0.035*
	(0.020)	(0.019)	(0.024)	(0.024)	(0.021)	(0.020)
Excluded F	931	951	950	927	1,172	1,145
Village observations	34,226	34,226	34,080	34,080	34,080	34080
Nearest factory FEs	Y	Y	Y	Y	Y	Y
Baseline geographic controls	Y	Y	Y	Y	Y	Y
Latitude-longitude polynomial	N	Y	N	Y	N	Y

Notes: Sample is a cross-section of all Indonesian villages within 100 km of a palm oil processor, excluding villages in cities and on Java. The treatment variable is set to 1 if a village is within 10 km of a palm factory and 0 otherwise. Changes in sample sizes are due to imperfect village matches across datasets. Baseline geographic controls include elevation, slope, historical precipitation, flow accumulation, distance to river, distance to major road in 2000, nighttime luminosity in 1993, distance to city, village area, and an urban dummy. A polynomial in latitude and longitude refers to the latitude and longitude of village centroids and their squared terms. Village definitions follow the 2016 BPS shapefile, and the nearest factory to every village is calculated in GIS using the 2016 Economic Census. Population is taken from the 2011 Village Census, firms from the 2016 Economic Census, and primary village income from the 2014 Village Census. Predicted probability is the predicted value for each village obtained from a probit model regressing the main suitability controls on the treatment. GAEZ palm suitability is the normalized value of the grid cell containing the village centroid. Relative suitability is the difference between the normalized potential palm oil yield and the average normalized yield of coffee, cocoa, and teas. Excluded F refers to the Kleibergen-Paap first-stage excluded F statistic; Montiel-Pflueger effective F statistics are similar. Robust standard errors are in parentheses and stars denote statistical significance at the 10 (*), 5(**), and 1(***) percent levels.

Table 6: Reduced-form Falsification Tests

Outcome	Population (n)		Firms present (=1)		Agriculture (=1)	
Column	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main estimation sample						
GAEZ palm suitability	160.557***	161.838***	0.036***	0.039***	-0.008*	-0.006
	(29.451)	(30.518)	(0.010)	(0.010)	(0.004)	(0.004)
Village observations	30,178	30,178	34,080	34,080	34,080	34080
Panel B: Restricted sample—zero first stage						
GAEZ palm suitability	14.876	9.213	0.013	0.018	-0.005	-0.004
	(31.714)	(33.657)	(0.011)	(0.011)	(0.005)	(0.005)
Village observations	15,516	15,516	17,934	17,934	17,934	17,934
Nearest factory FEs	Y	Y	Y	Y	Y	Y
Baseline geographic controls	Y	Y	Y	Y	Y	Y
Latitude-longitude polynomial	N	Y	N	Y	N	Y

Notes: Full sample is a cross-section of all Indonesian villages within 100 km of a palm oil processor, excluding villages in cities and on Java. Restricted sample additionally excludes villages within 25 km of a factory. Baseline geographic controls include elevation, slope, historical precipitation, flow accumulation, distance to river, distance to major road in 2000, nighttime luminosity in 1993, distance to city, village area, and an urban dummy. A polynomial in latitude and longitude refers to the latitude and longitude of village centroids and their squared terms. Village definitions follow the 2016 BPS shapefile. Population is taken from the 2011 Village Census, firms from the 2016 Economic Census, and primary village income from the 2014 Village Census. GAEZ palm suitability is the normalized value of the grid cell containing the village centroid. Robust standard errors are in parentheses and stars denote statistical significance at the 10 (*), 5(**), and 1(***) percent levels.

Table 7: Processor Density and External Economies of Scale

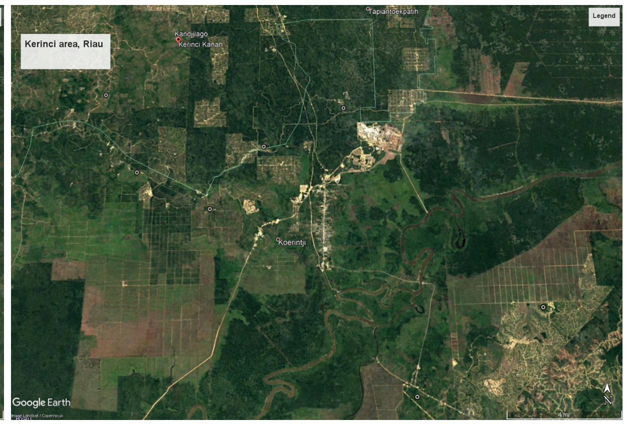
Potential buyer buffer zone	25 km		50 km		25 km		50 km	
Column	1	2	3	4	5	6	7	8
Panel A Outcome	**Primary source of income is agriculture (=1)**				**Population (n)**			
Potential buyers (n)	-0.003***	-0.000	-0.001***	-0.000	49.554***	13.768**	10.274***	1.986
	(0.001)	(0.001)	(0.000)	(0.000)	(6.762)	(6.848)	(1.704)	(1.938)
Near factory treatment (=1)		-0.012**		-0.018***		239.818***		324.487***
		(0.005)		(0.005)		(42.784)		(42.390)
Potential buyers X treatment		-0.003***		-0.001*		39.058***		8.406***
		(0.001)		(0.000)		(7.718)		(2.418)
Observations	34,226	34,226	34,226	34,226	30,288	30,288	30,288	30,288
Panel B outcome	**Firms (n)**				**Firms (=1)**			
Potential buyers (n)	0.086***	0.033**	0.011***	-0.000	0.007***	0.002	0.002***	0.001
	(0.015)	(0.014)	(0.004)	(0.005)	(0.001)	(0.001)	(0.000)	(0.000)
Near factory treatment (=1)		0.831***		1.062***		0.129***		0.139***
		(0.136)		(0.143)		(0.009)		(0.009)
Potential buyers X treatment		0.036**		-0.001		0.001		-0.000
		(0.016)		(0.005)		(0.001)		(0.000)
Observations	34,226	34,226	34,226	34,226	34,226	34,226	34,226	34,226

Notes: Sample is a cross section of Indonesian palm oil farmers observed in the farmer cost survey fielded with the 2013 Agricultural Census. The presence of a factory in a subdistrict and the number of potential buyers at a 50 kilometer radius is calculated in GIS, and merged to the household survey. Changes in sample size are due to imperfect merges, missing data, farmers not using urea, and logging zero. District fixed effects and additional controls for farm size, household size, number of household members, sex, age, and education attainment included throughout. Cluster-robust standard errors are in parentheses and stars denote statistical significance at the 10 (*), 5(**), and 1(***) percent levels.

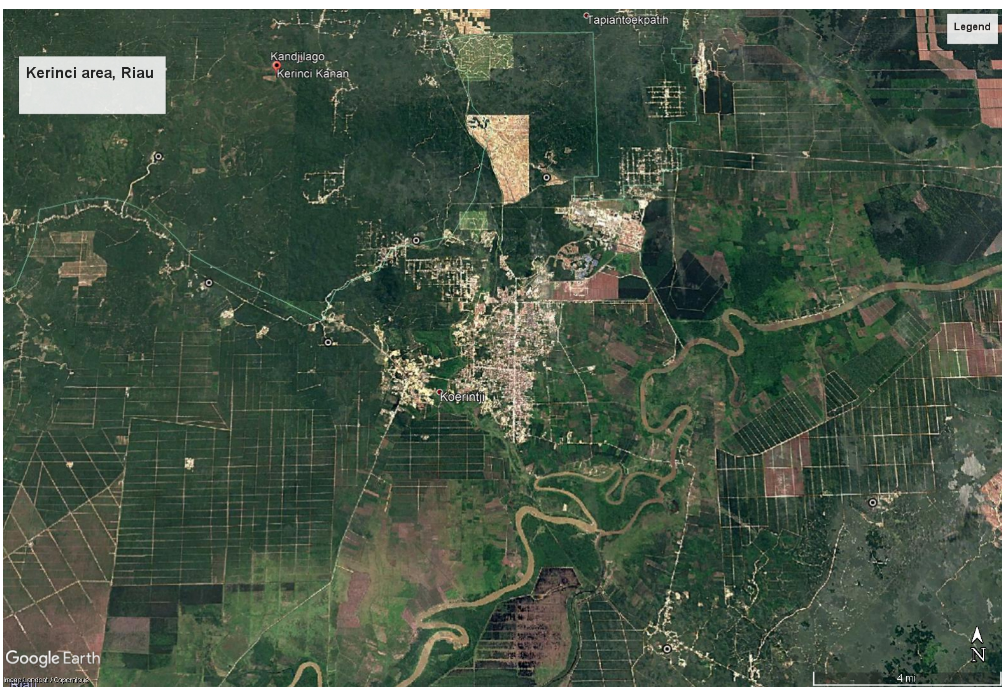
Figure 1: Kerinici Area, Riau—LandSat since 1984

(a) 1984

(b) 2000

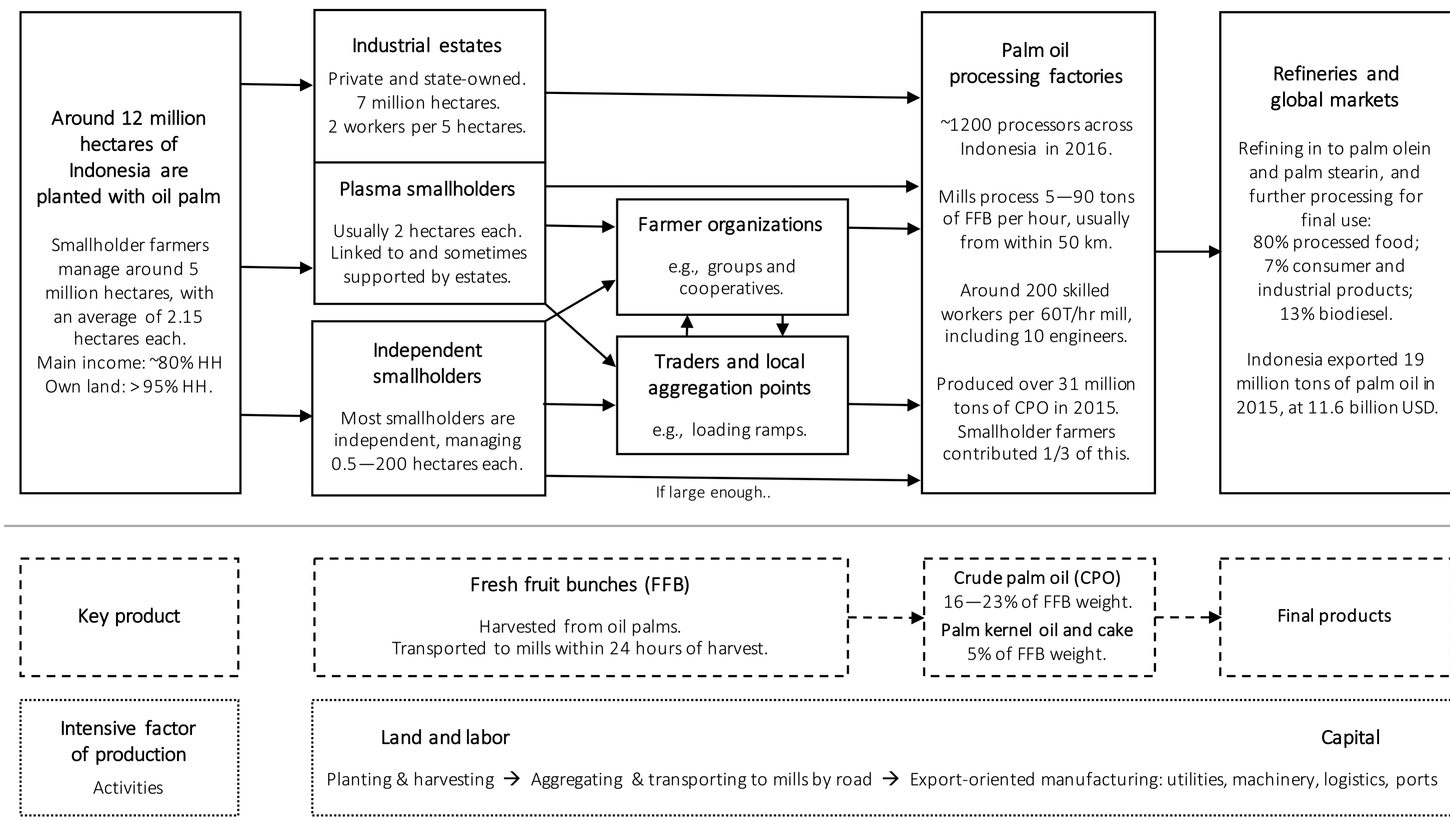


(c) 2016



Notes: This figure shows the long-term change in local economic activity with the emergence of palm oil processing factories in the Kerinci area of Riau province. Landsat imagery is directly as screenshots from Google Earth Pro's desktop historical slider. Modern palm oil processors in 2016 are the black dots in each image.

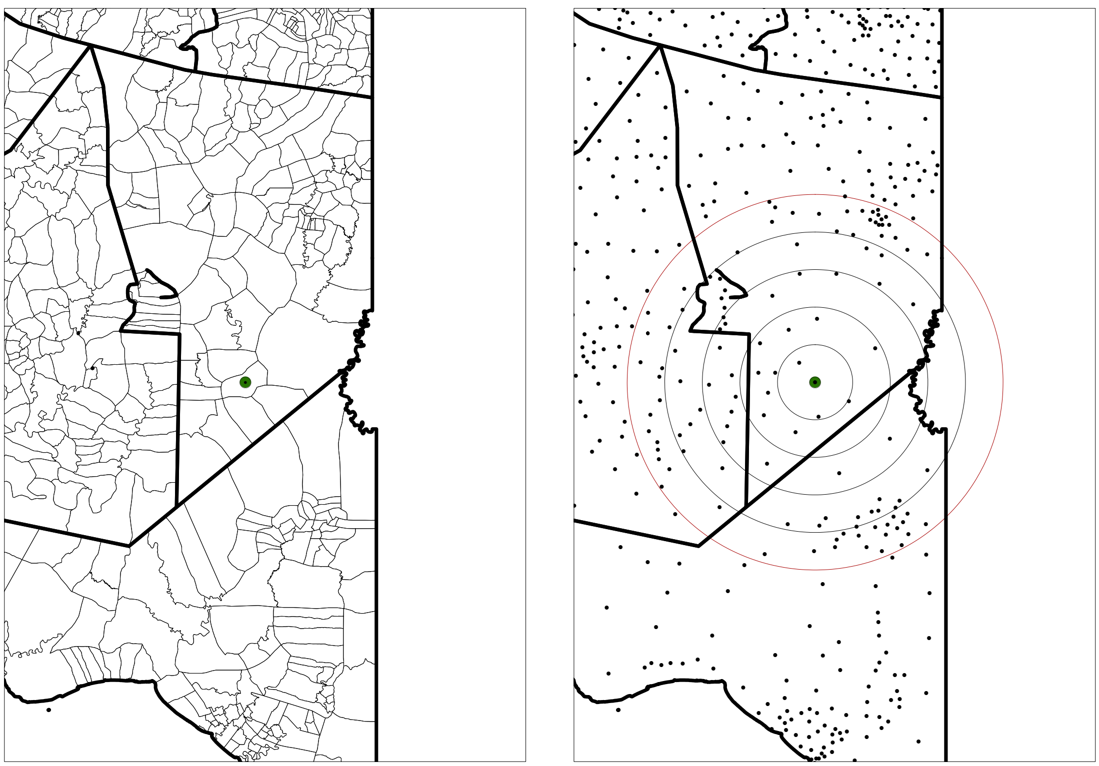
Figure 2: The Farm-to-Factory Supply Chain



Notes: Author's own depiction. Figures are for Indonesia, from no earlier than 2013, and sourced from official government statistics, site visits, personal discussions, and correspondence.

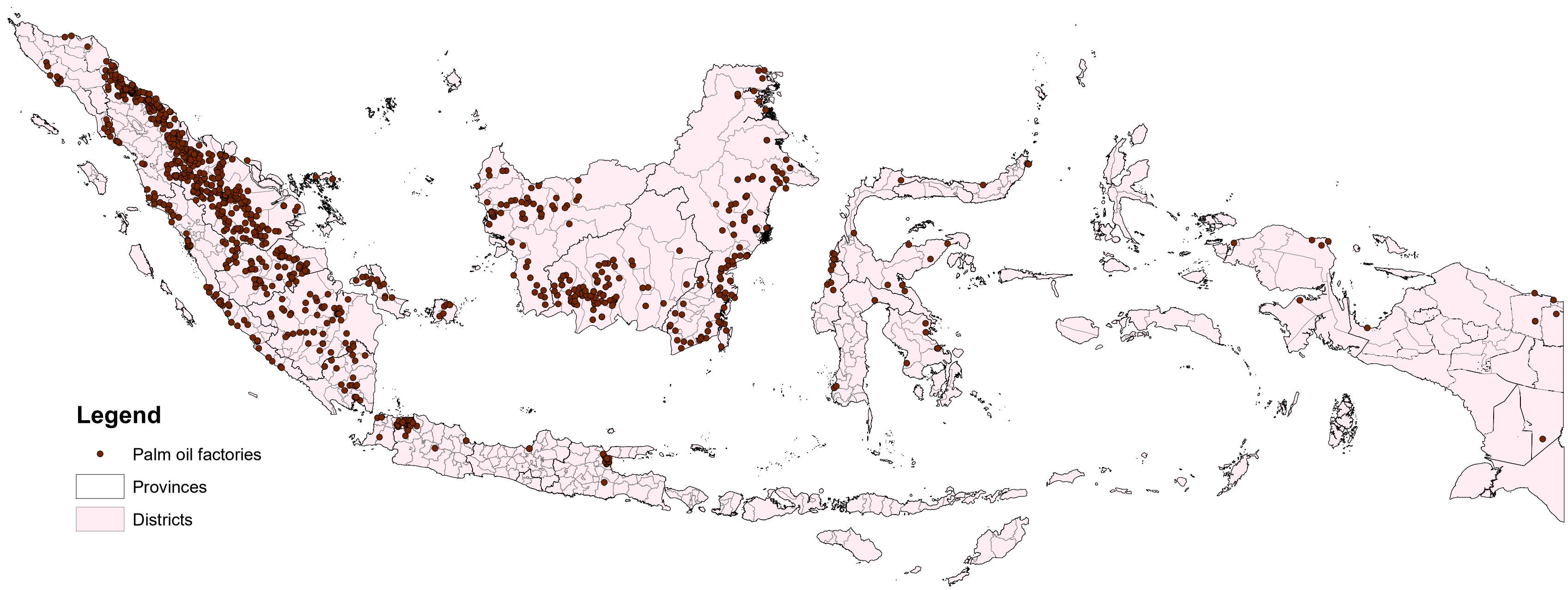
Figure 3: Research Design Intuition

(a) Villages　　　　(b) Centroids



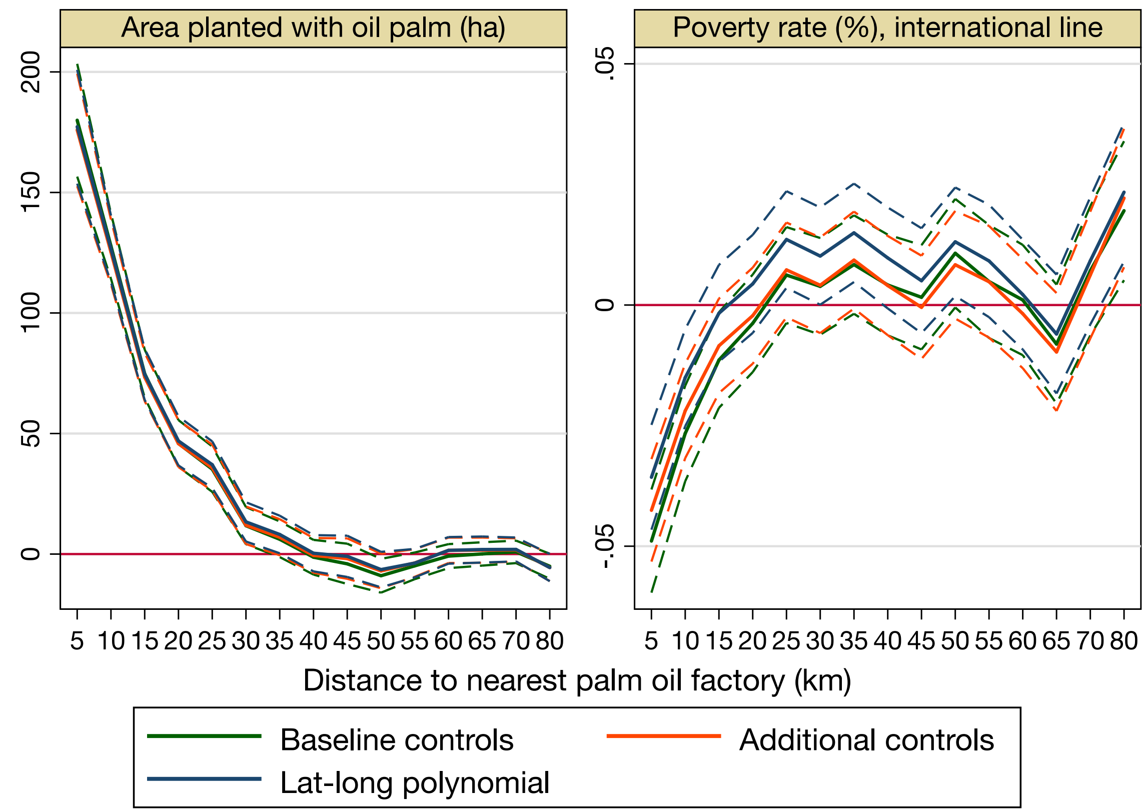
Notes: This figures show the intuition of my research design with the mapping of village polygons (grey lines in A) to village centroids (black dots in B), and the overlay of buffers around a factory (green dot) in Papua province. Dark lines are district boundaries. Each buffer is 20 km farther away from the factory. Beyond the red 100 km buffer village observations are discarded. Please note that (a) all estimates in the article use 5 km intervals, and (b) Papua is a region where villages tend to be larger and factories sparse, allowing a clearer illustration here.

Figure 4: Indonesian Palm Oil Processors, 2016



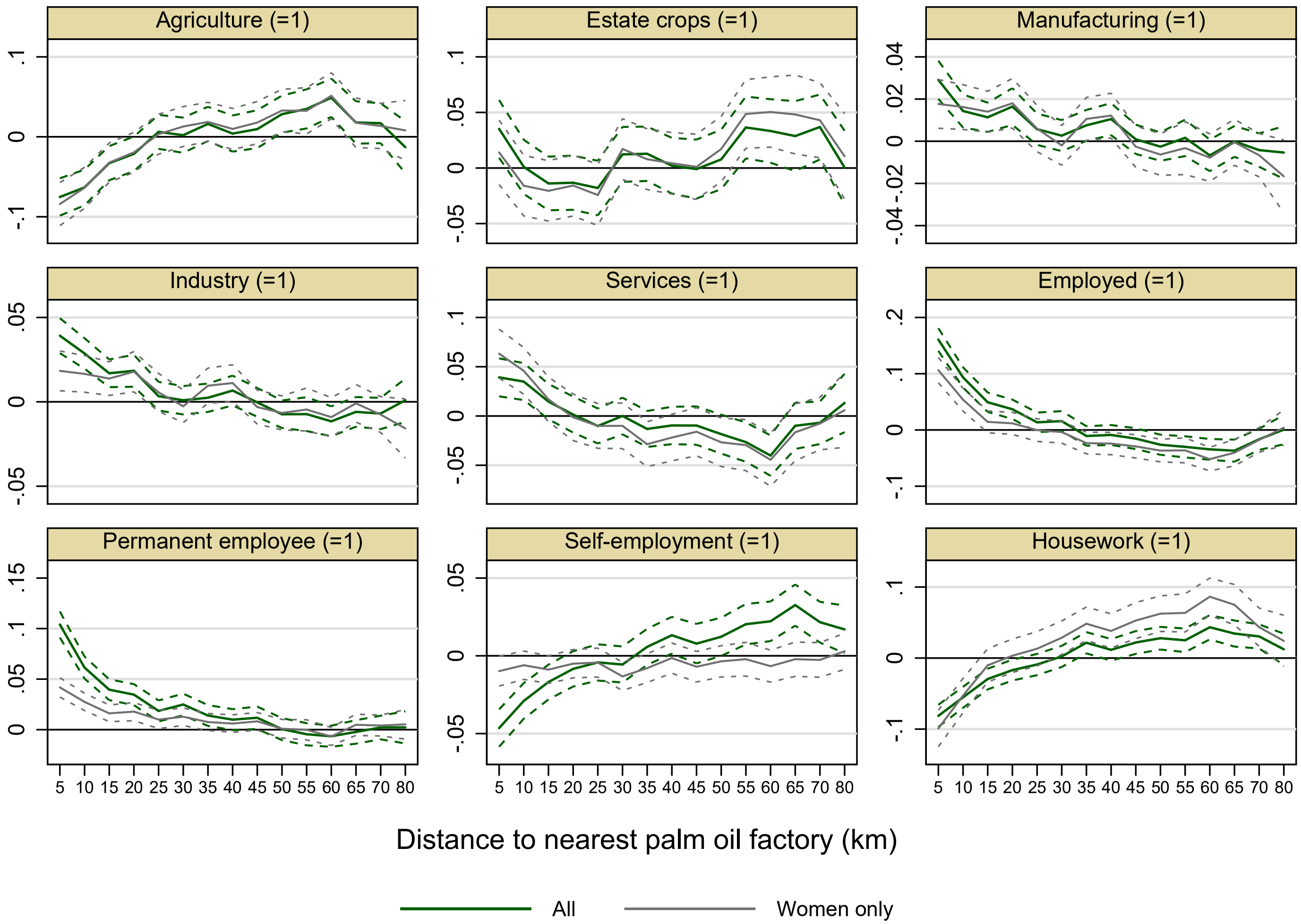
Notes: This figure plots the locations of all palm processing firms reported in the 2016 Economic Census of Indonesia. Firms are assigned the centroid of the village where domiciled. Grey lines are districts and dark lines are provinces as defined in the 2010 Population Census shapefile.

Figure 5: Proximate Adoption and Poverty



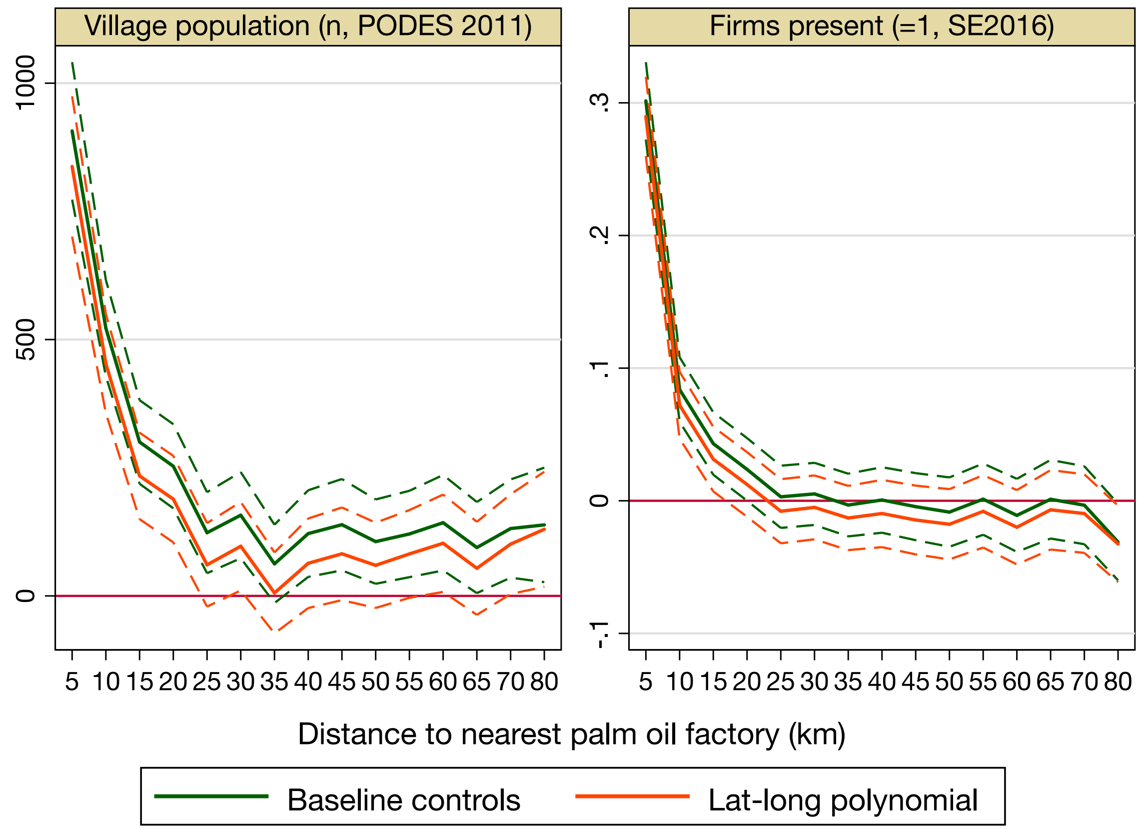
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using the village oil palm acreage and poverty rates measured at the international line as dependent variables, showing an increase in area planted with palm oil and decrease in estimated poverty closer to palm factories. 95% confidence intervals are represented by the dotted lines. Additional controls add to the baseline controls travel time to the nearest city, travel cost to the nearest city, and indicators for river, coast, plains, valleys, forest proximity (in, near, and outside) and forest function. Lat-long polynomial further adds a polynomial in latitude and longitude. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure 6: Local Labor Markets



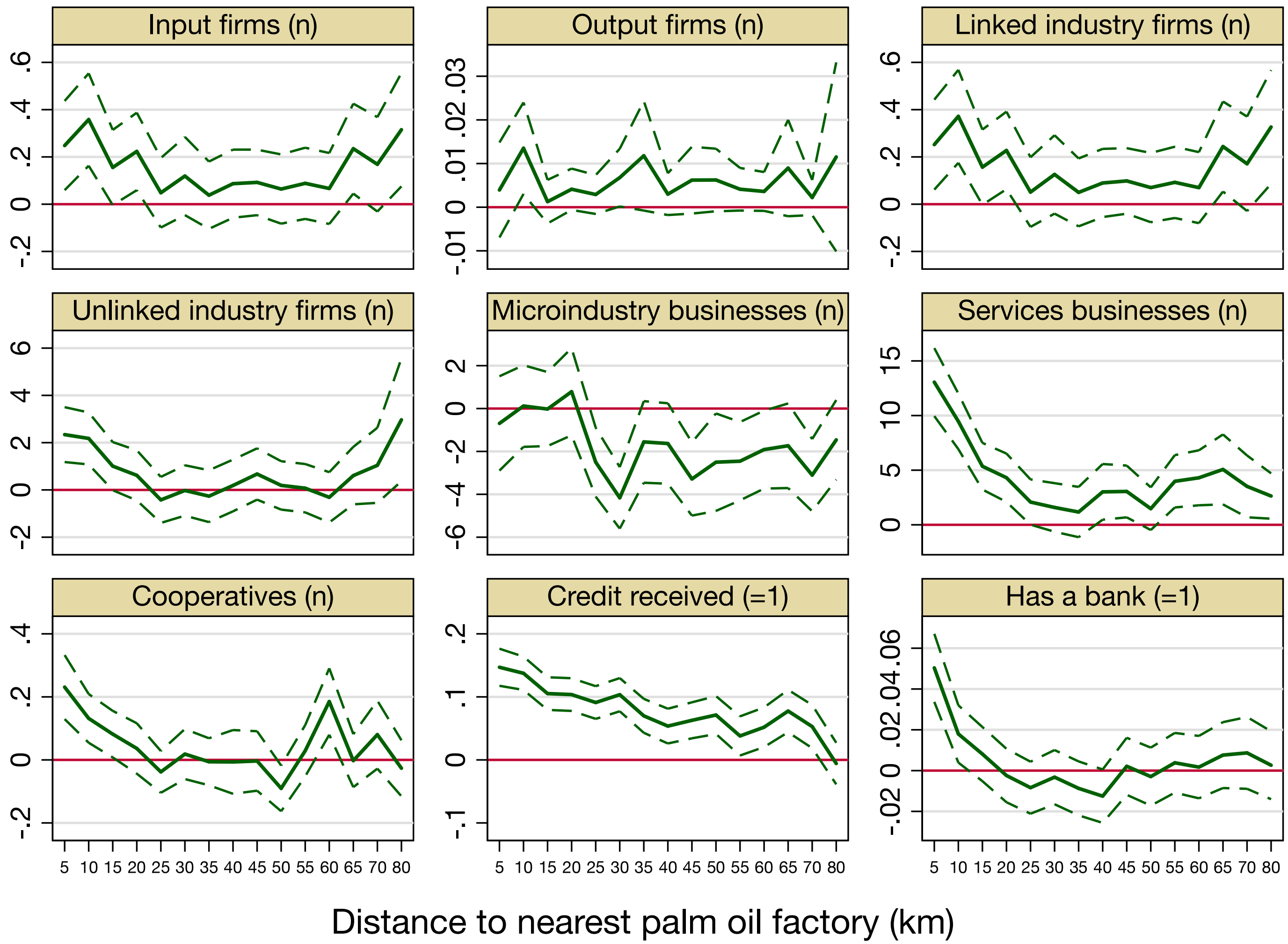
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable the main sector of employment and employment status at the individual level as observed in the 2010 Population Census (10 percent random sample, stratified on village by the author). 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Robust standard errors are clustered on village. Note the different axis scales when visually interpreting.

Figure 7: Population and Firms



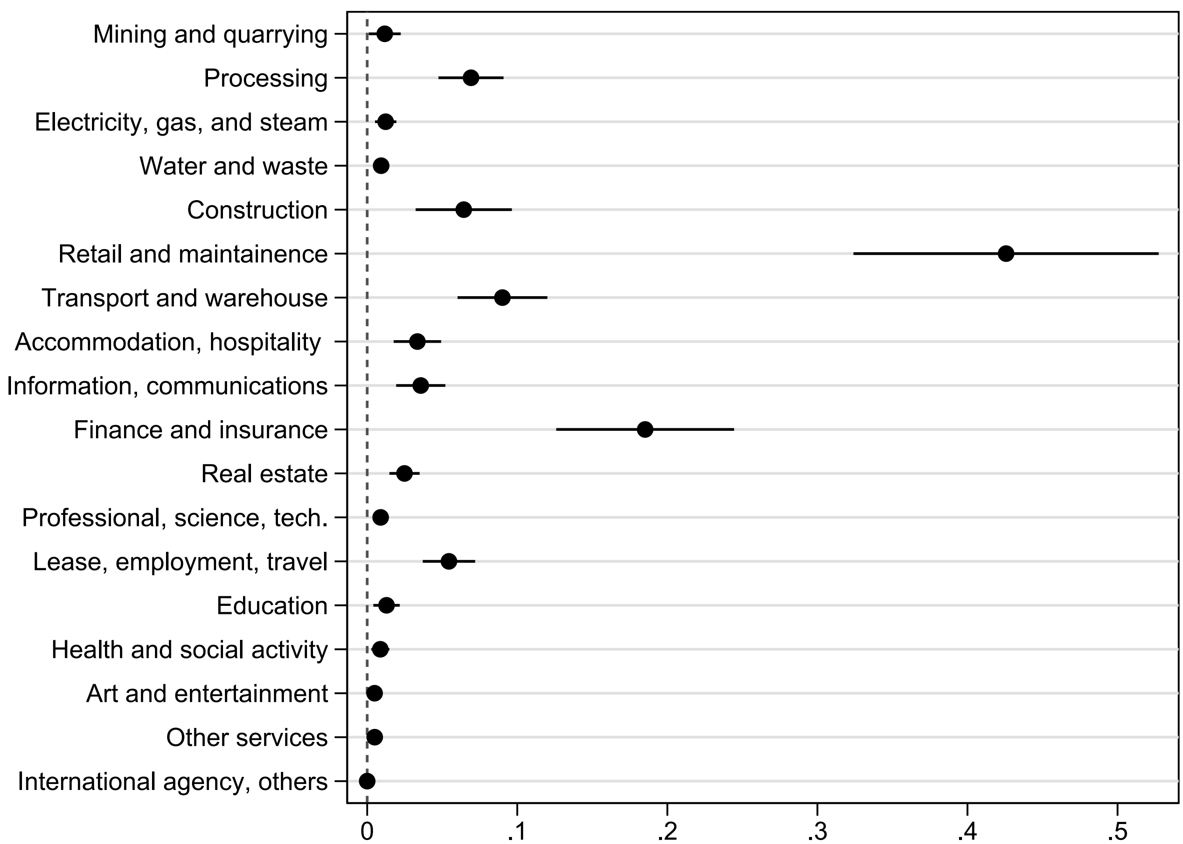
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable village population measured in the 2011 Village Census and the whether a village has a firm present according to the 2016 Economic Census. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure 8: Production and Consumption Linkages



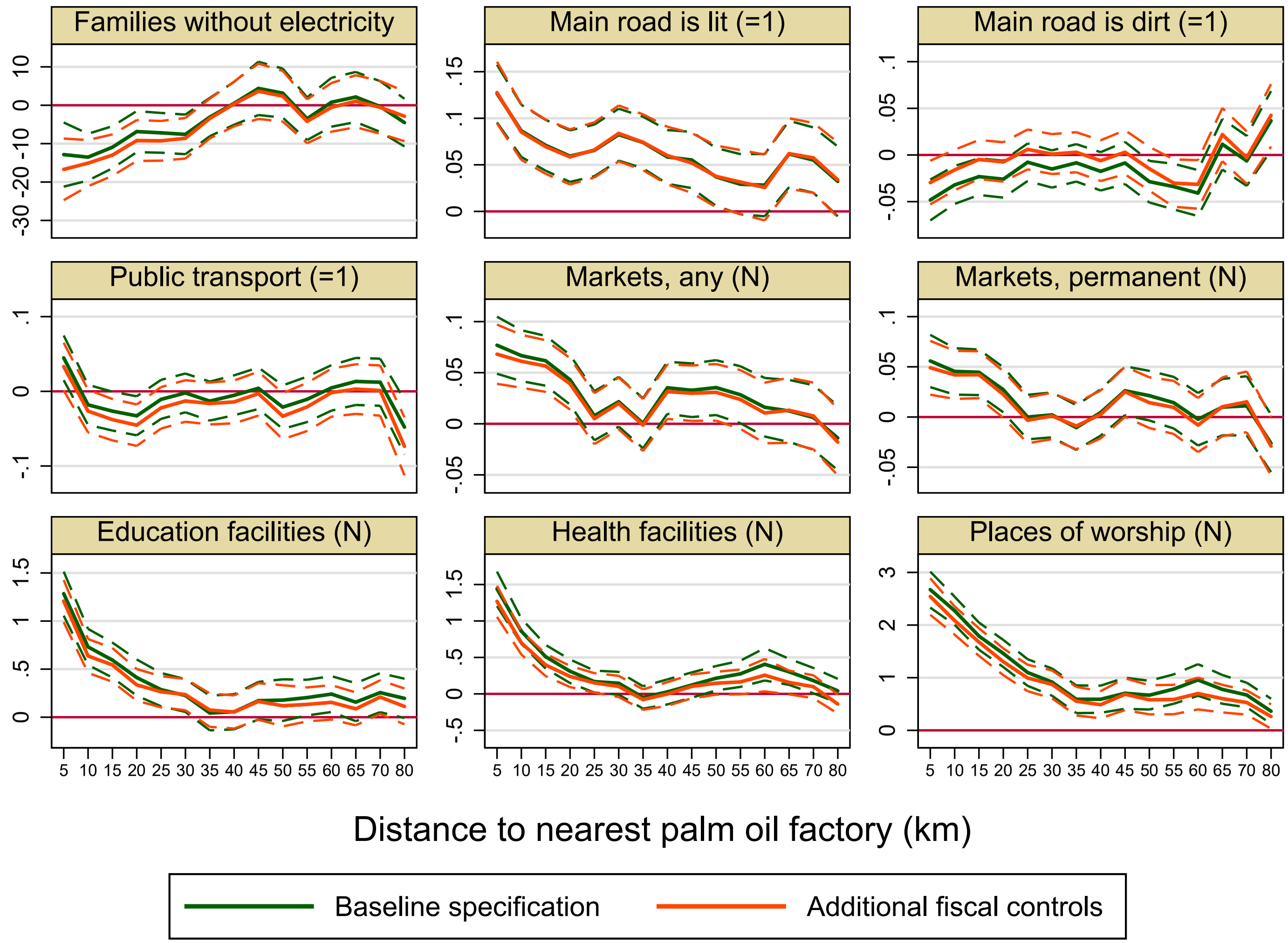
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable the number of firms in each village, as measured in the 2016 Economic Census (formal firms) and the 2014 Village Census. Input-output linkages (for formal firms in Economic Census) are identified through the 2010 Input-Output Table. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Note the different axis scales when visually interpreting.

Figure 9: Proximate Firms By Sector



Notes: This figure plots treatment effects from estimating equation one using as the treatment a 10 km from a factory dichotomous treatment indicator, and using as dependent variables the number of firms by industry. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure 10: Village Public Goods, 2014 Village Census



Notes: These figures plot the coefficients from estimating Equation 1 at 5km bins using as a dependent variable different public good outcomes reported in the 2014 Village Census. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. The green lines indicate distance band coefficients from my baseline specification, while the red lines indicate those obtained from the same regressions adding village revenue and expenditures as control variables.

Online Appendix—Not For Publication

Appendix A—Data

Overview

Dataset	Years	Source	Variables
Agricultural Census	2013	BPS	area planted by crop
Economic Census	2016	BPS	number of firms by type, palm oil processors
Population Census	2010	BPS	employment by sector employment status migration status
Village Census (PODES)	1993– 2014	BPS	population, organizations, distance to city & capital, other village characteristics
SUSENAS	2005– 2013	BPS	household consumption, sector of employment
Poverty map	2015	SMERU	village poverty estimates
Hydrosheds–WWF	N/A	USGS	slope, flow accumulation, distance to river, elevation
CHIRPS	2018	UCSB	historical precipitation
Road map	2000	GoI	distance to road
World Port Index	2015	WPI	distance to port
DMSP	1992	NOAA	night time luminosity
Global Forest Watch	2019	WRI	palm oil mills, date-stamped by hand
Global Agroecological Zones	N/A	FAO	agro-climatic suitability

Detailed Dataset and Variable Descriptions

Agricultural Census, 2013. *Badan Pusat Statistik (BPS).* A census of all agricultural households, conducted every 10 years, which I use to calculate the area planted in each village by crop.

Economic Census, 2016. *BPS.* Census of all firms, conducted every 10 years, which I use to calculate the number of firms by size and sector in each village. The 2016 Economic Census is also used to identify all palm oil processing firms, which are then geolocated to the centroid of the village in which they are domiciled, as observed in the official 2016 BPS village map, using an exact matching procedure on village codes. The nearest processors, the distance to the nearest processor, and the number of possible buyers for fruit for each village are then calculated in ArcMap. Together with the BPS 2010 Input-Output Table, the KBLI 2015 codes assigned to observations in the Economic Census are also used to count the number of forward and backward linked firms in each village. This procedure is described and a list of linked industries provided in Appendix B.

Population Census, 2010. *BPS.* The Population Census is the principal way to measure Indonesia's population, demographics, and migration extent. Migration is assessed through questions on whether people live in different district or provinces to that which they lived in 5 years ago or at birth. The census also reports employment by sector and status, which I combine with migration information to examine employment effects by migration status. I draw a ten percent random sample from the complete census, stratified on village, to ease computation.

Village Census, 1993–2014. *BPS.* The census of village heads, Potensi Desa (PODES), is conducted roughly every three years for most all villages (desa) and urban wards (kelurahan) throughout Indonesia. I draw several key outcomes and controls from the village census, for example related to population, village organizations, public goods (schools, health clinics, and churches/mosques), and the primary source of village income. I also draw several key controls, including the distance and travel time to the nearest city and a host of geographical identifiers, for example related to coast proximity, forest proximity, whether a villages is on a slope, plain, or valley, whether a village has a river, and more.

National Socioeconomic Survey, 2005–2013. *BPS.* The National Socioeconomic Survey SUSENAS is Indonesia's main at-least-annual, district-representative household survey, covering over 2 million households in its core wave each year. Because SUSENAS is only representative at the district level, I pool observations over the years for which I have village identifiers to improve village coverage. The primary goal of incorporating SUSENAS is to move beyond the sector of employment (which SUSENAS has) examine expenditures (i.e., consumption).

Indonesia Poverty Map, 2015 *SMERU Research Institute.* The SMERU Research Institute produces a poverty map of Indonesia, which using populations censuses, SUSENAS, and standard poverty mapping techniques (e.g., like those used at the World Bank) to estimate poverty (P0, P1, P2) at Indonesia's lowest administrative level, the village. The SMERU Poverty Map is available from SMERU's website and on request from their research team.

Hydrosheds-WWF *USGS.* The Hydrosheds-WWF online data portal provides extremely fine geospatial data related to geography. Using their various raster and shapefile products and the 2016 BPS village map, I calculate in ArcMap for every village its slope, water flow accumulation, the distance to the nearest river, and elevation.

Major road map, 2000. *Government of Indonesia.* I use a map of Indonesia's major roads in 2000 to calculate in ArcMap the distance to the nearest major road from each 2016 village.

World Port Index, 2015. *WPI.* The World Port Index provides a shapefile geolocating ports all around the world. Using ArcMap and the 2016 village map, I calculate the distance from each village to its nearest port.

Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS), 1992–2015. *National Oceanic and Atmospheric Administration (NOAA), U.S. Government.* I calculate from the Nighttime Lights Time Series Version 4, Defense Meteorological Program Operational Linescan System the average nighttime luminosity for each village. Calculations are done in ArcMap by assigning village centroids the values of the centroids of their nearest raster grid cell. The goal of bringing this data into my analysis is to allow me to capture pre-period local economic development and industry. Using the gridded data and 2016 village maps allows me to preserve all villages in my dataset. By contrast, linking villages back across administrative datasets, which also contain such proxies, is more challenging to match villages over time.

Global Agro-Ecological Zones. *Food and Agriculture Organization of the United Nations.* GAEZ uses agronomic models and high resolution geographic and climatic data to predict attainable yields for different crops on each piece of land regardless of whether the land is cultivated. It does not rely on actual cultivation in its estimates, nor does it involve estimating any sort of statistical relationship between observed inputs, outputs, and agro-climatic conditions. By excluding investments in productivity that might be considered outcomes, GAEZ offers plausibly exogenous variation in crop-specific geographic endowments by construction. The benefit of GAEZ data over using the plausibly exogenous model inputs (e.g., rain, soil type, temperature) is the additional crop-specific yield responses which, without the agronomic model, these variables do directly speak to. Fischer, van Nelthuizen, Shah, and Nachtergaele (2002) detail GAEZ construction. Costinot, Donaldson, and Smith (2016) and Nunn and Qian (2011) discuss additional benefits of GAEZ for identification. Village potential agro-climatically attainable yields for different crops are calculated in ArcMap by assigning village centroids the values of the centroids of their nearest raster grid cell. Note that the level spatial aggregation in the GAEZ data is large than many villages, giving limited variation within a particular locality.

Global Forest Watch Palm Oil Mills. *World Resources Institute.* The Global Forest Watch palm oil mills dataset, while counting less mills than official sources, exactly locates mill coordinates. I use this dataset to visually inspect the historical LandSat satellite imagery and identify establishment years. Due to cloud cover and poor image quality before 2004, establishment years for many factories could not be identified. I use this dataset to identify a subset of "new" mills built since 2000 and conduct placebo tests on outcomes in 1993.

Appendix B—Industries Linked to Palm Oil

Input—51/181 sectors (37th in value)

Oil Palm, Other Seasonal Plantations, Other Annual Plantations, Biopharmaca plants, Livestock and their results except fresh milk, Wood, Agricultural Services, Forestry Support Services, Fisheries Support Services, Tapestry, Rope and Other Textiles, Industrial Preservation and Tanning of Leather, Paper and Cardboard, Industrial Paper and Cardboard Goods, Manufacture of Printed Goods, Basic Chemical Industries Except Fertilizers, Fertilizers, Pesticides, Manufacture of Oil Refinery Products, Plastic Products, Magnetic Media and Optical Media, Glass Industry and Glass Products, Industrial Kitchen Tools, Carpentry and Agriculture from Metals, Hand-Mobilized Hand Tool Industry, Other Metal Goods Industry, Machinery and Equipment Industry, Other Equipment Repair Services, Other Industrial Items, Electricity, Natural and Artificial Gas, Steam / Hot Water Supply, Cold Air and Ice Production, Procurement of Water, Waste and Recycling Processing, Building Construction, Special Construction, Agricultural Infrastructure, Roads, Bridges and Ports, Wholesale and Retail Trade in addition to Cars and Motorcycles, Car and Motorcycle Trade, Car and Motorcycle Repair, Restaurants, Hospitality, Railway Transport, Highway Transportation, Sea Freight, River and Lake Transport, Air Freight, Transportation Support Services, Information Services, Communication, banks, Insurance and Pension Funds, Financial Support Services, and Company Services.

Output—9/181 sectors

Oil Palm, Animal Oil and Vegetable Oil , Peeled Grain, Animal Feed, Pesticide, Drugs, Pharmaceutical Products, Soap and Cleaning Products, Industrial Cosmetic Products, and Other Chemical Goods.

Notes

Linked industries are identified through BPS' 2010 Input-Output Table. Concordance to the KBLI 2015 codes in the 2016 Economics Census was done for each 2010 linked sector by hand, translating then matching codes on descriptions. All output sectors were matched successfully. 9/51 input sectors were not: (1) other seasonal plantations, (2) agricultural services, (3) industrial paper and cardboard goods, (4) industrial kitchen tools, carpentry, and agriculture, (5) hand mobilized hand tool industry, (6) other industrial items, (7) agricultural infrastructure, (8) financial support services, and (9) company services. Agricultural services, tools, agricultural infrastructure, and financial and company services are likely relatively important; estimates without them should be interpreted as such.

Appendix C—Supplementary Figures

Identification Checks

Additional Results

Additional Robustness Checks

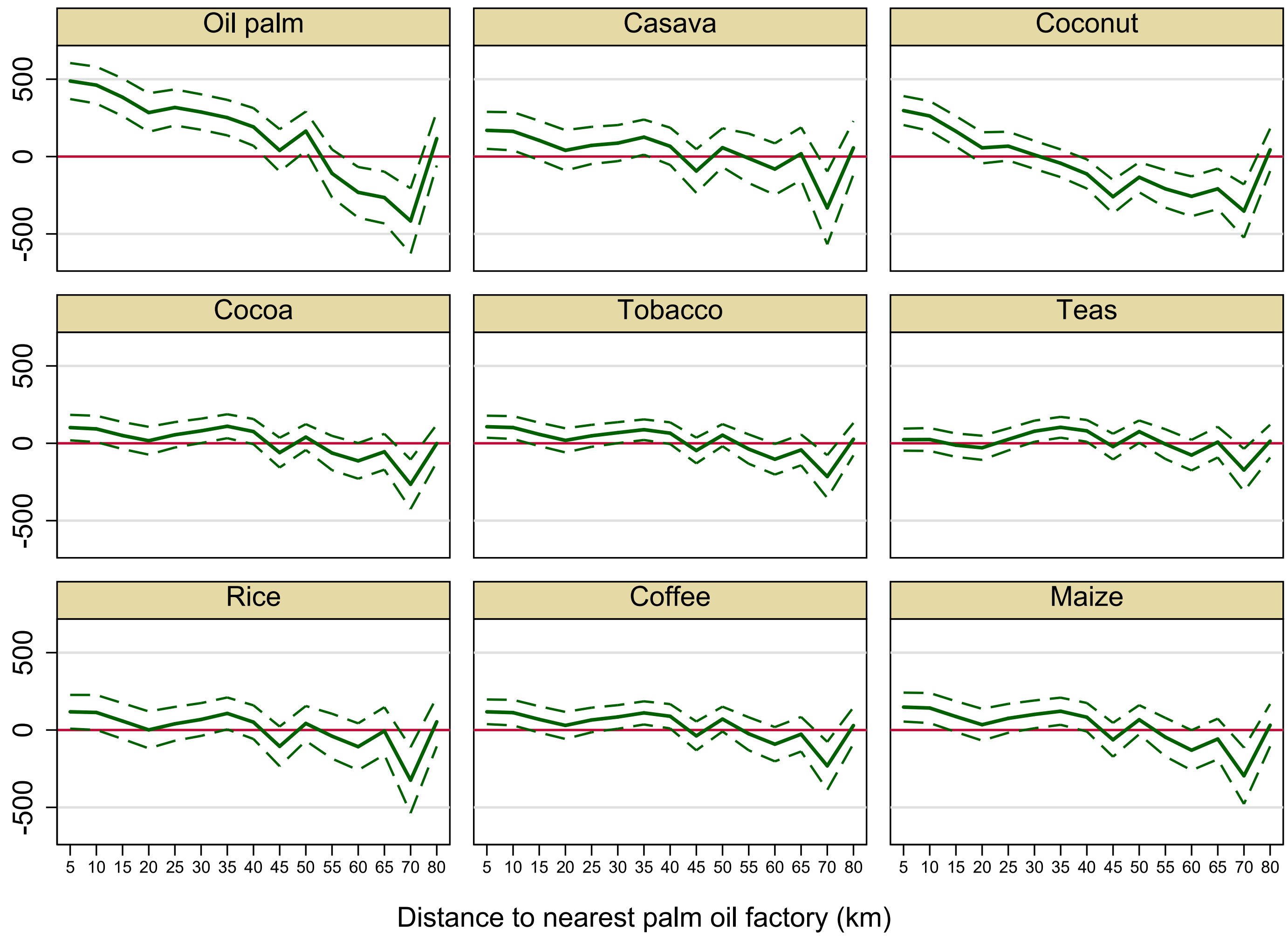
No sample restrictions, all villages

No nearest factory fixed effects, only covariates and district FEs

No covariates, only nearest factory FEs

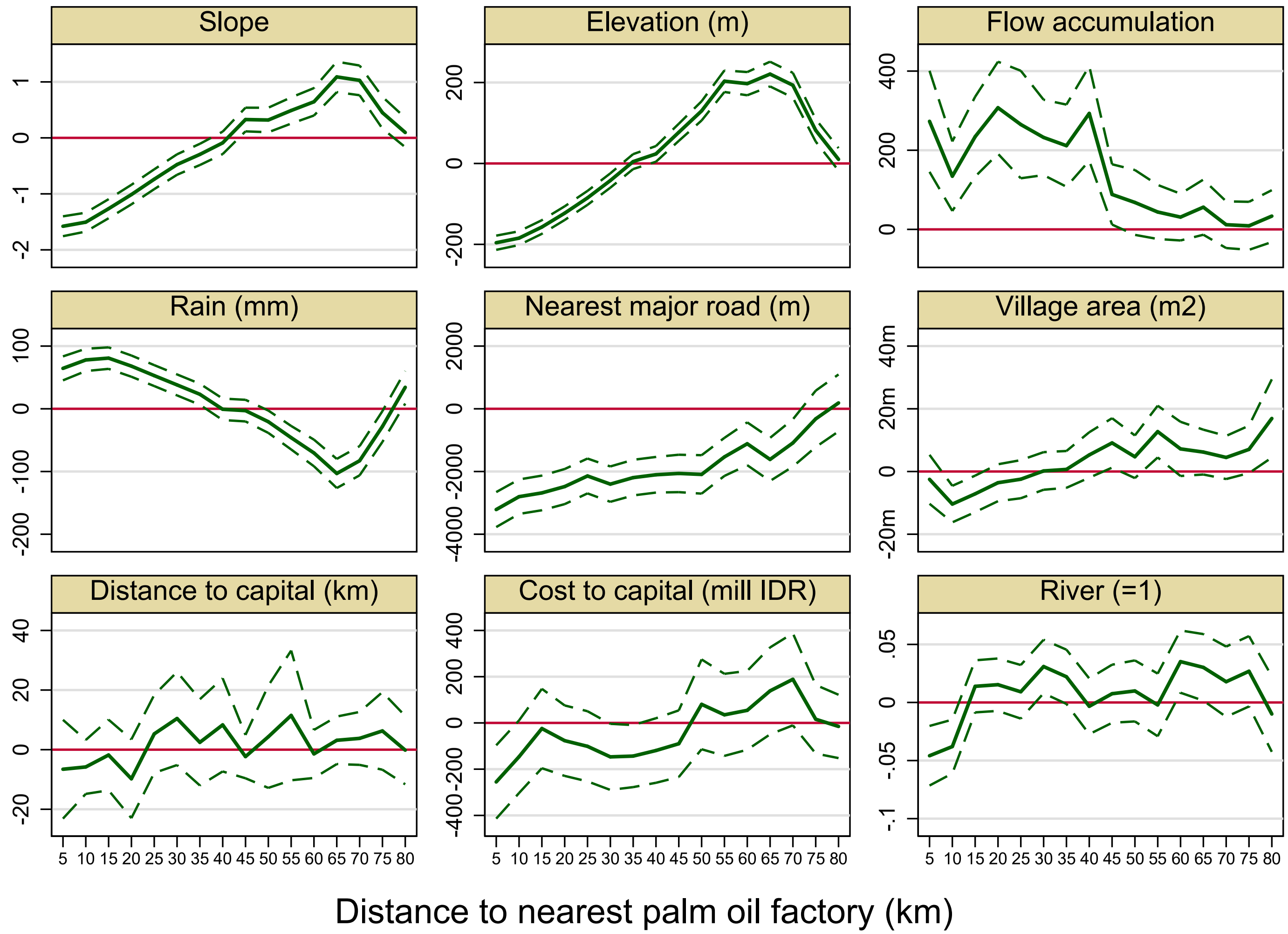
Additional Descriptive Figures

Figure A1: Selection on FAO-GAEZ Agro-Climatic Suitability



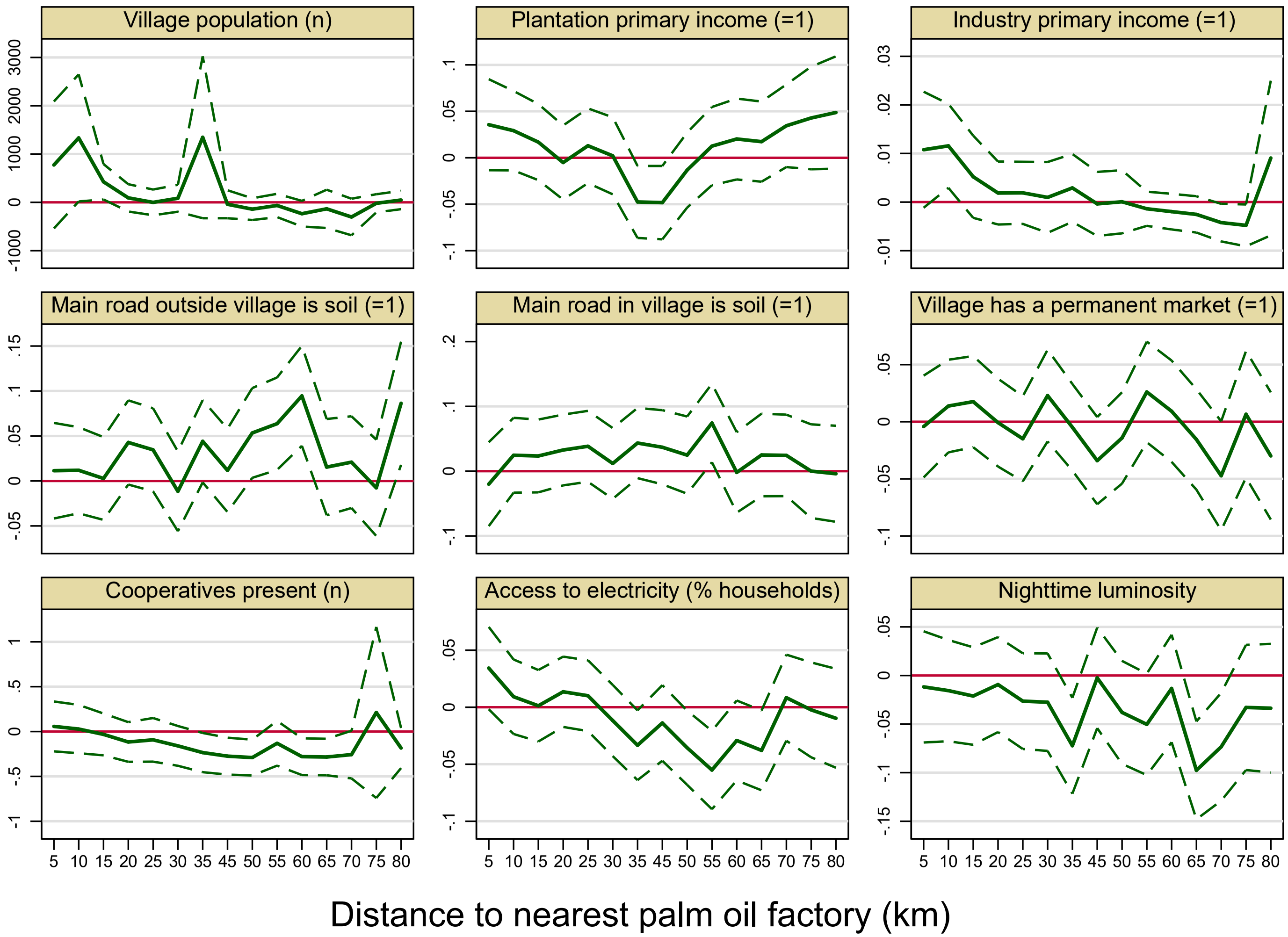
Notes: These figures show for factory locations are most closely related to suitability for oil palm and not other crops, with the exception of coconut which is also a palm tree. The figures plot the coefficients from estimating Equation 1 at 5km bins using as dependent variables the agro-climatically attainable yields, for each crop, of the overlapping or nearest gridcell for each village. Confidence intervals are for 95%.

Figure A2: Geographic Selection and Balance



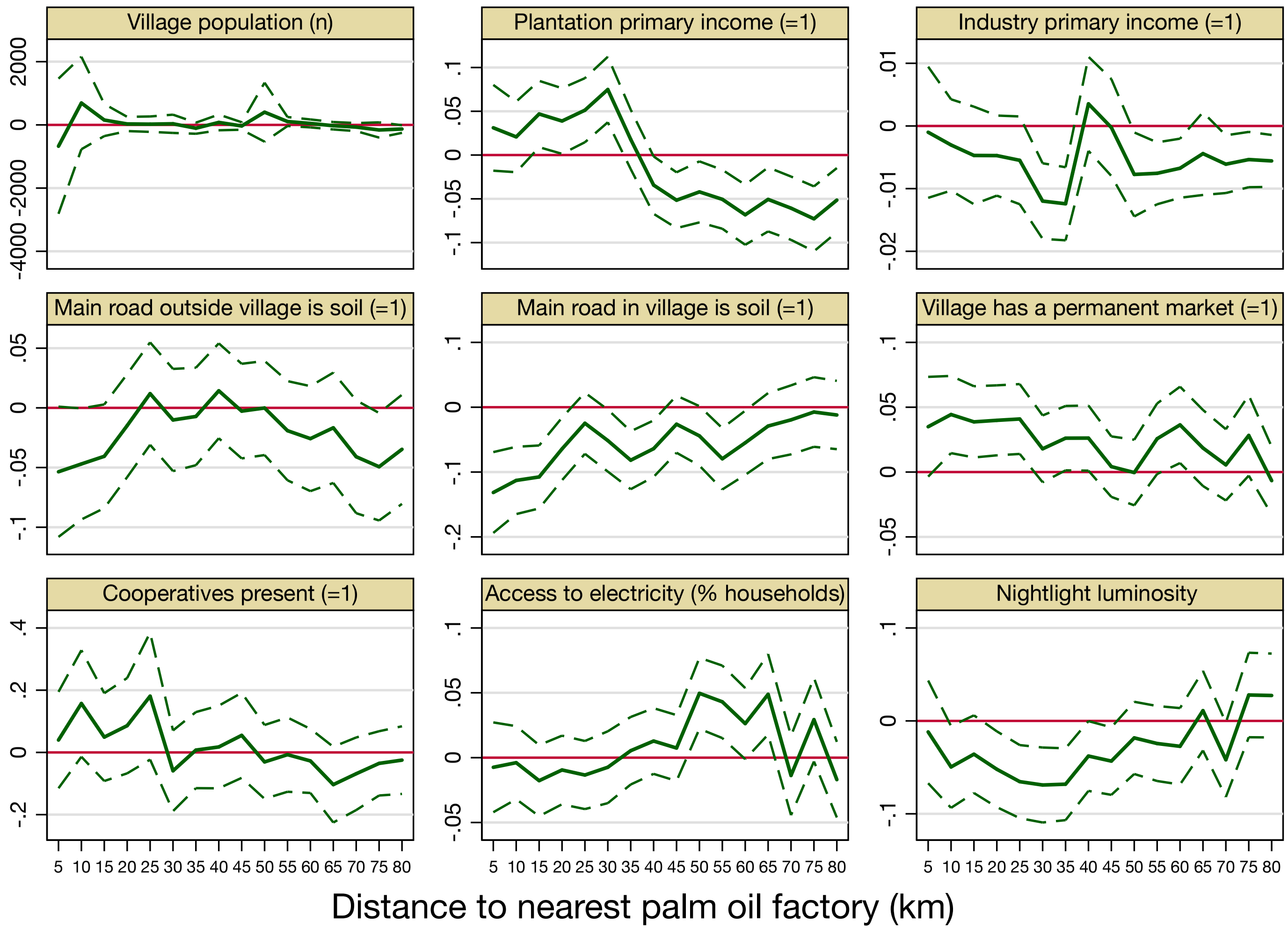
Notes: These figures show how geographic conditions more closely related to growing conditions, are quite different near factories, but other geographic characteristics show less of a relationship. The figures plot the coefficients from estimating Equation 1 at 5km bins using geographic variables calculated in GIS or taken from PODES 2014 as the dependent variable. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.95% confidence intervals are represented by the dotted lines.

Figure A3: Pre-period Placebo Tests—Districts with No Palm Production in 2000



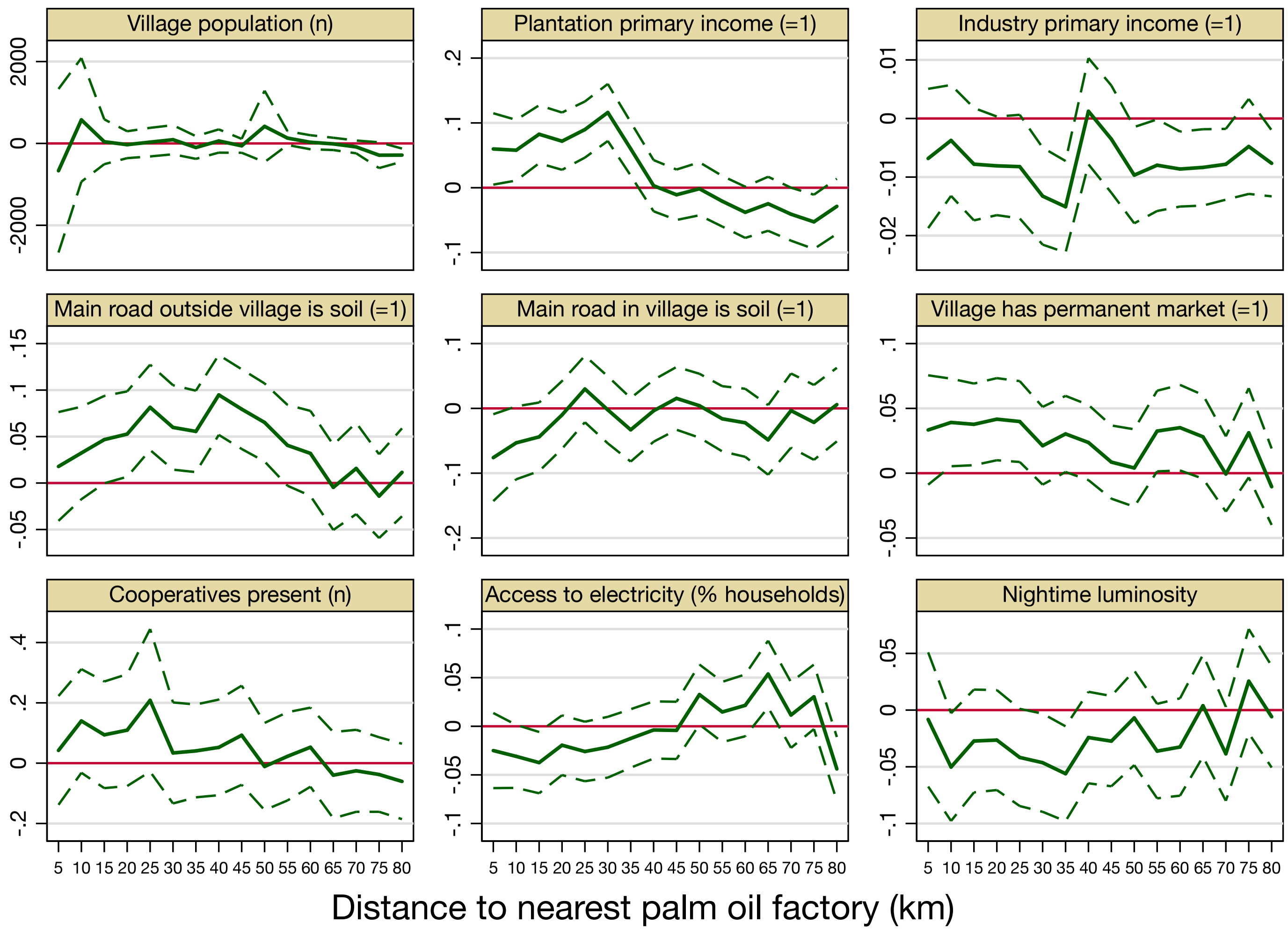
Notes: These figures show how the main patterns documented in the paper are far less clear in the historical data before the factories existed. The figures plot the coefficients from estimating Equation 1 at 5km bins using geographic variables calculated in GIS or taken from PODES 2014 as the dependent variable. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.95% confidence intervals are represented by the dotted lines.

Figure A4: Pre-period Placebo Tests—New Factories Since 2000



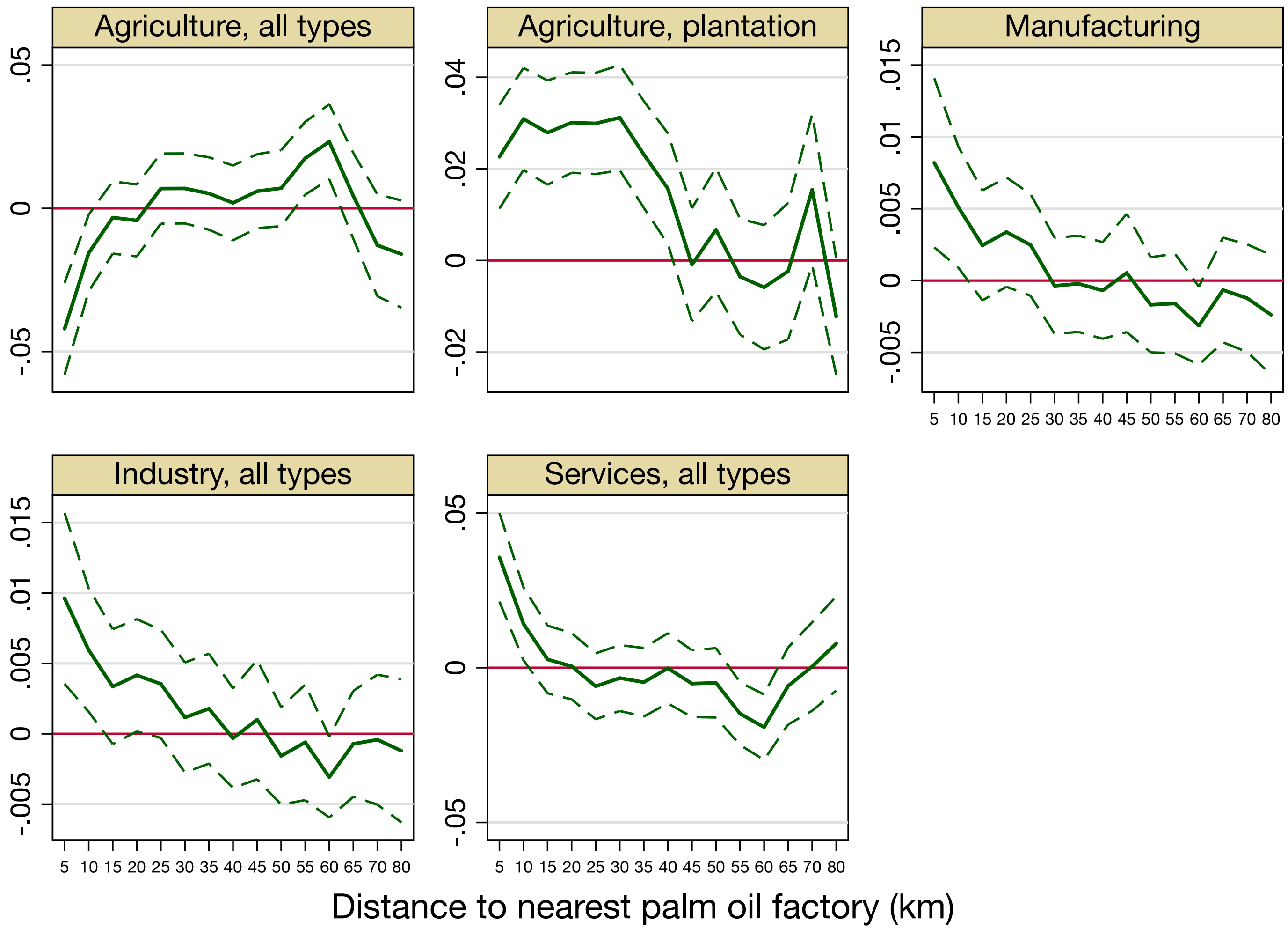
Notes: These figures show how the main patterns documented in the paper are far less clear in the historical data before the factories existed. The figures plot the coefficients from estimating Equation 1 at 5km bins using geographic variables calculated in GIS or taken from PODES 2014 as the dependent variable. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.95% confidence intervals are represented by the dotted lines.

FIGURE A5: PRE-PERIOD PLACEBO TESTS—NEW FACTORIES SINCE 2010



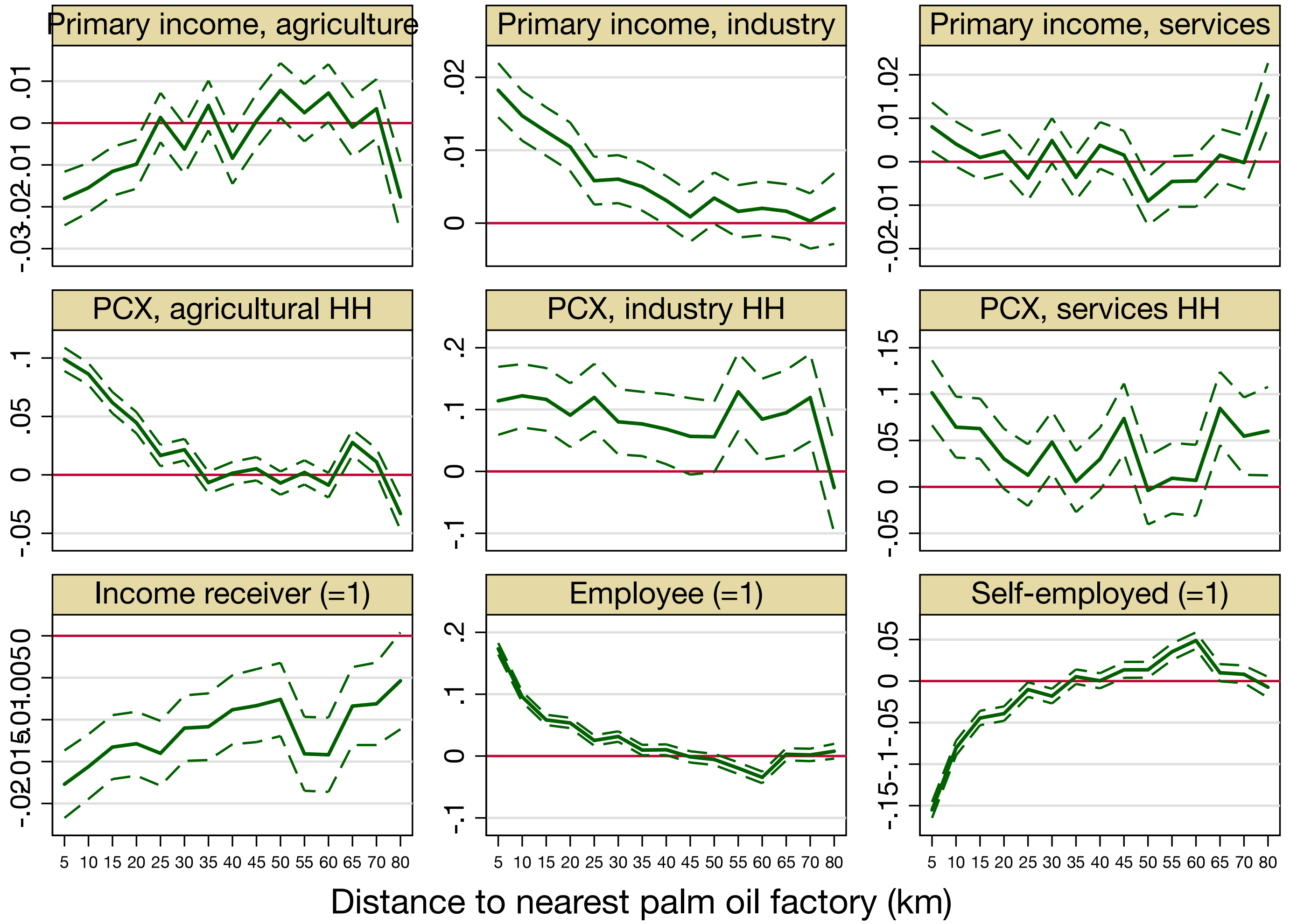
Notes: The figures plot the coefficients from estimating Equation 1 at 5km bins using geographic variables calculated in GIS or taken from PODES 2014 as the dependent variable. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.95% confidence intervals are represented by the dotted lines.

Figure A6: Primary Sector of Village Income—2014 Village Census



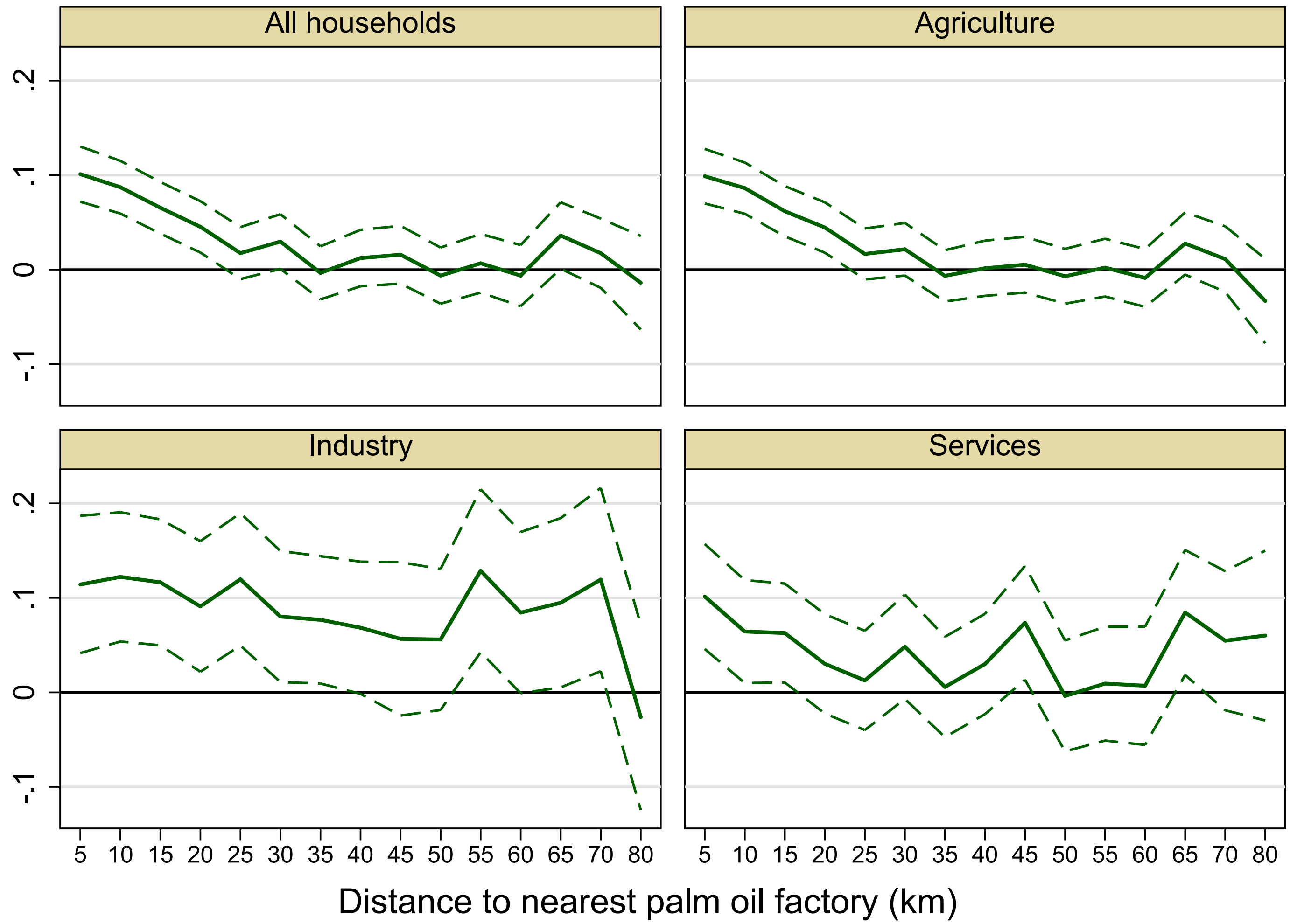
Notes: These figures show local labour market patterns as proxied by the primary source of income and livelihoods in villages as reported by village heads in teh village censuses. The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable the main source of income for the village, as measured in PODES 2014. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A7: Employment, Expenditure, and Work Status—Pooled SUSENAS



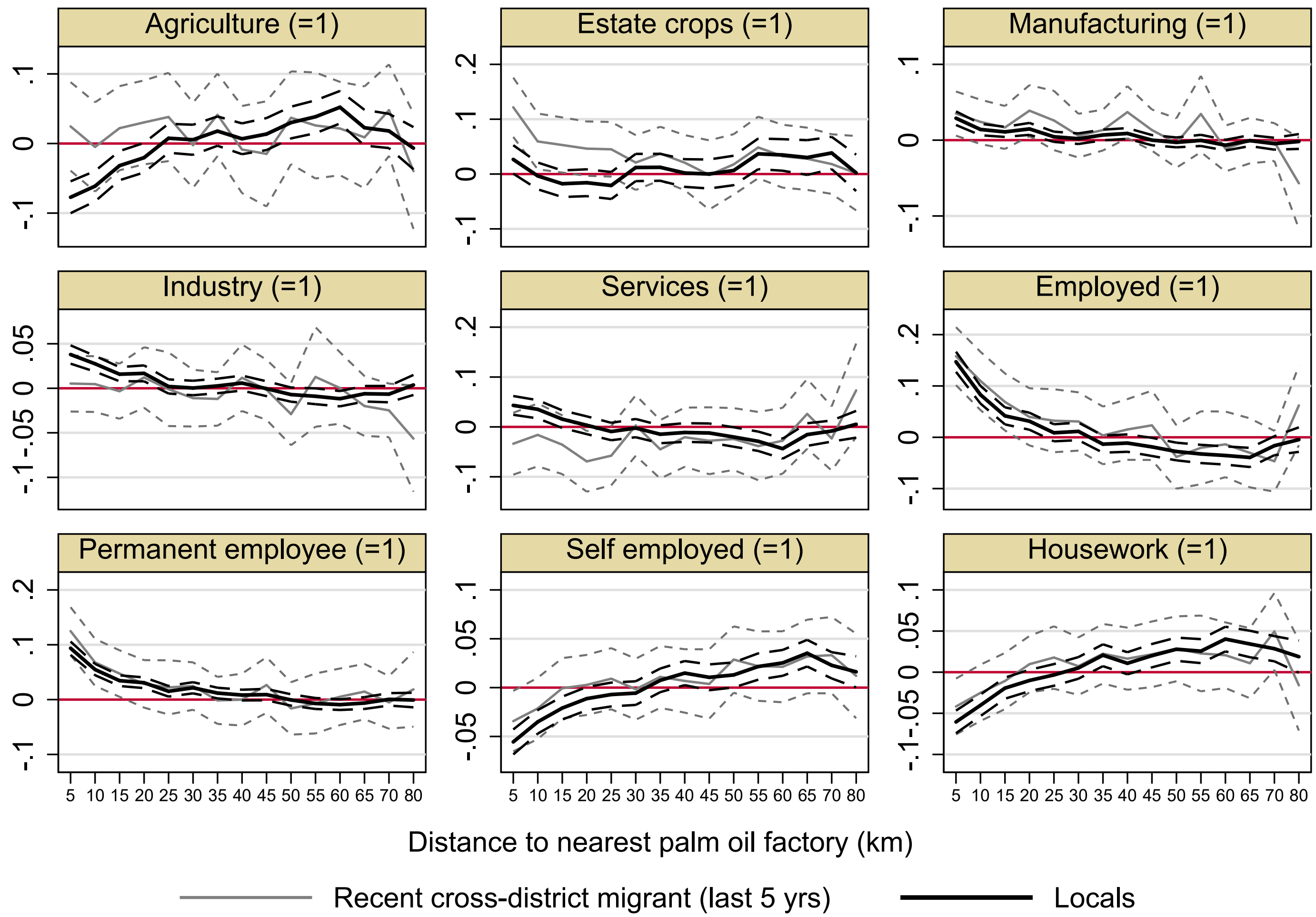
Notes: These figures show employment, expenditure, and work status effects using the national socioeconomic survey SUSENAS, pooled over time as it does not cover every village in every year. The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable per capita household expenditures, work status, and the probability of deriving most of your household income from a particular sector, as measured in SUSENAS pooled from 2006–2011. Pooling increases village coverage as the survey is designed to only be representative at the district level. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A8: Household Consumption—Pooled SUSENAS



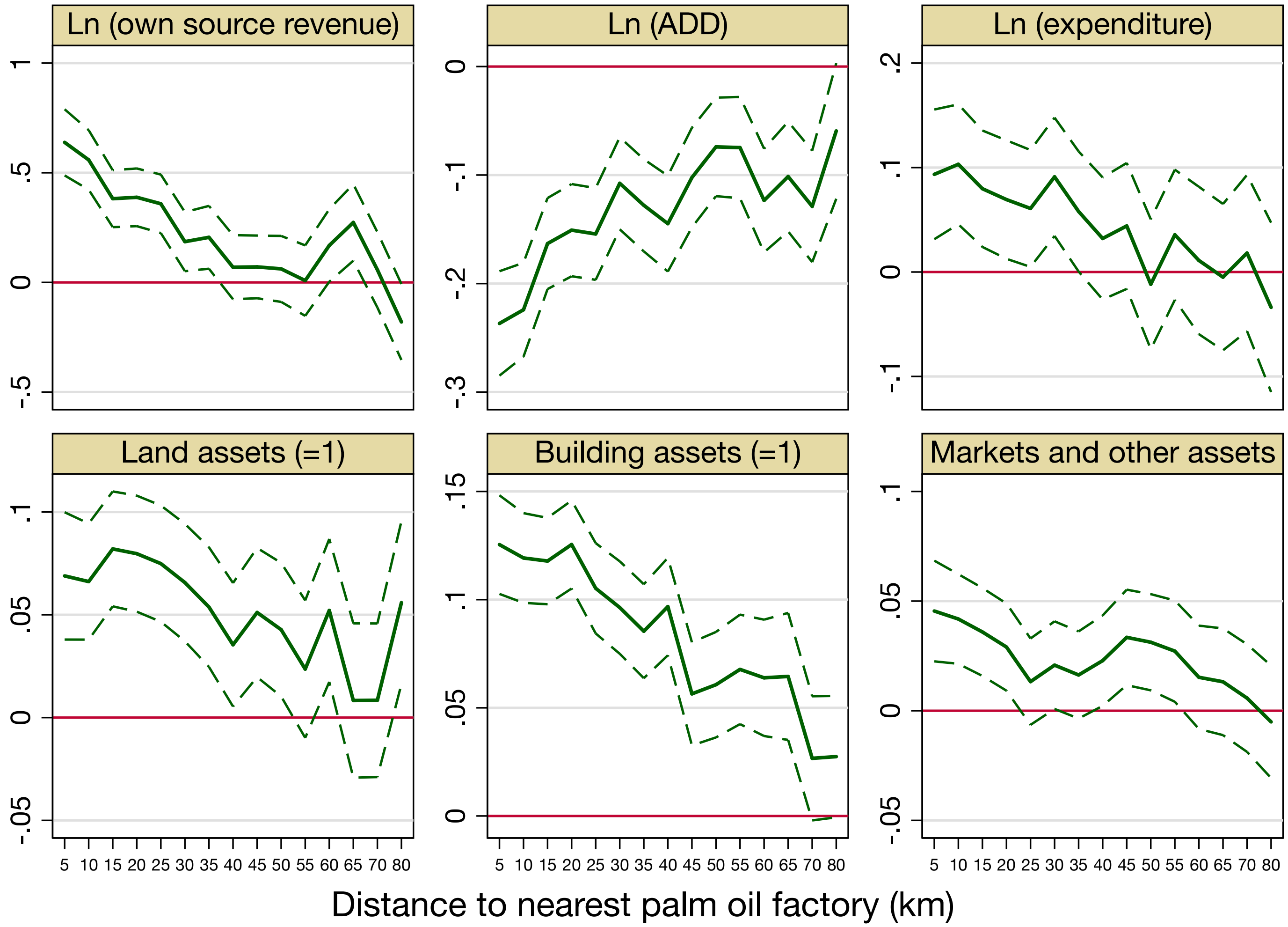
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable per capita household expenditures reported in the national socioeconomic survey SUSENAS, pooled across all years for which village identifier codes are available. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Robust standard errors are clustered on village, and survey-year fixed effects are include throughout.

FIGURE A9: EMPLOYMENT IMPACTS BY MIGRATION STATUS—2010 POPULATION CENSUS



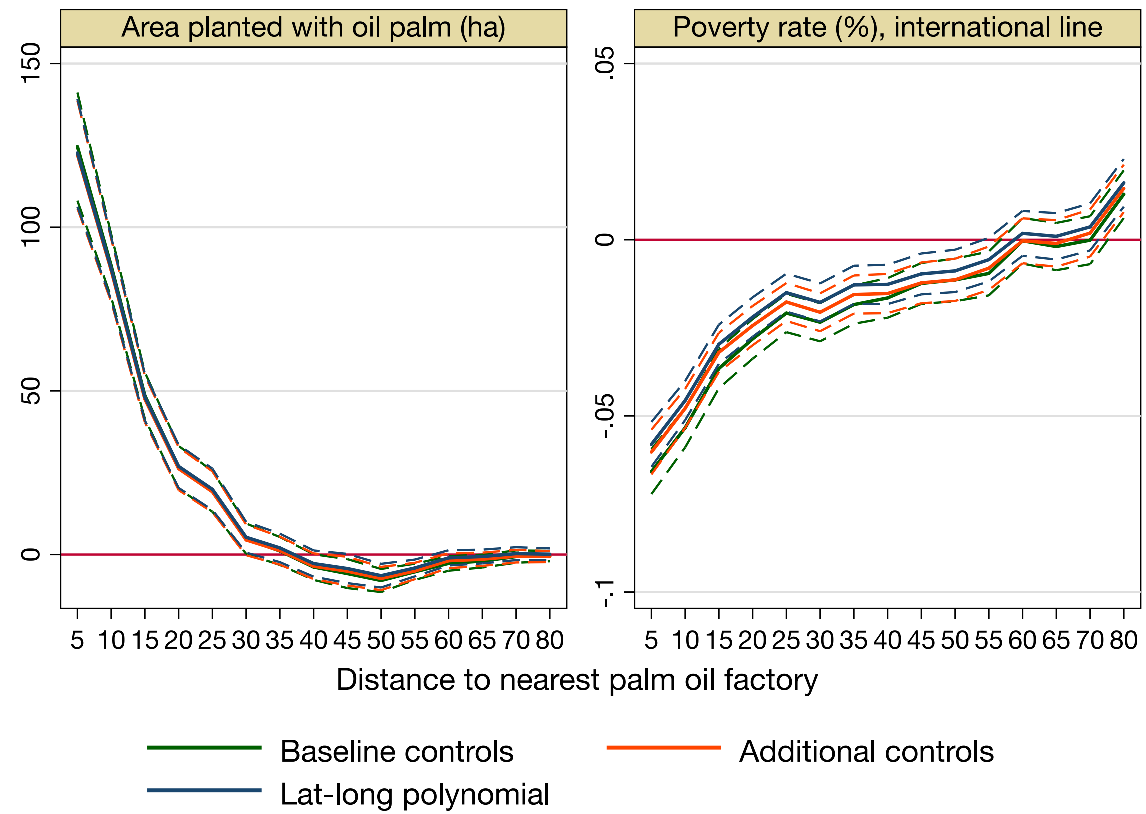
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable employment sector and status, as measured in the 2010 Population Census. Samples are split by "locals" (orange) and people who lived in a different district 5 years ago (purple). 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

FIGURE A10: IMPACTS ON VILLAGE FINANCES AND ASSETS—2014 VILLAGE CENSUS



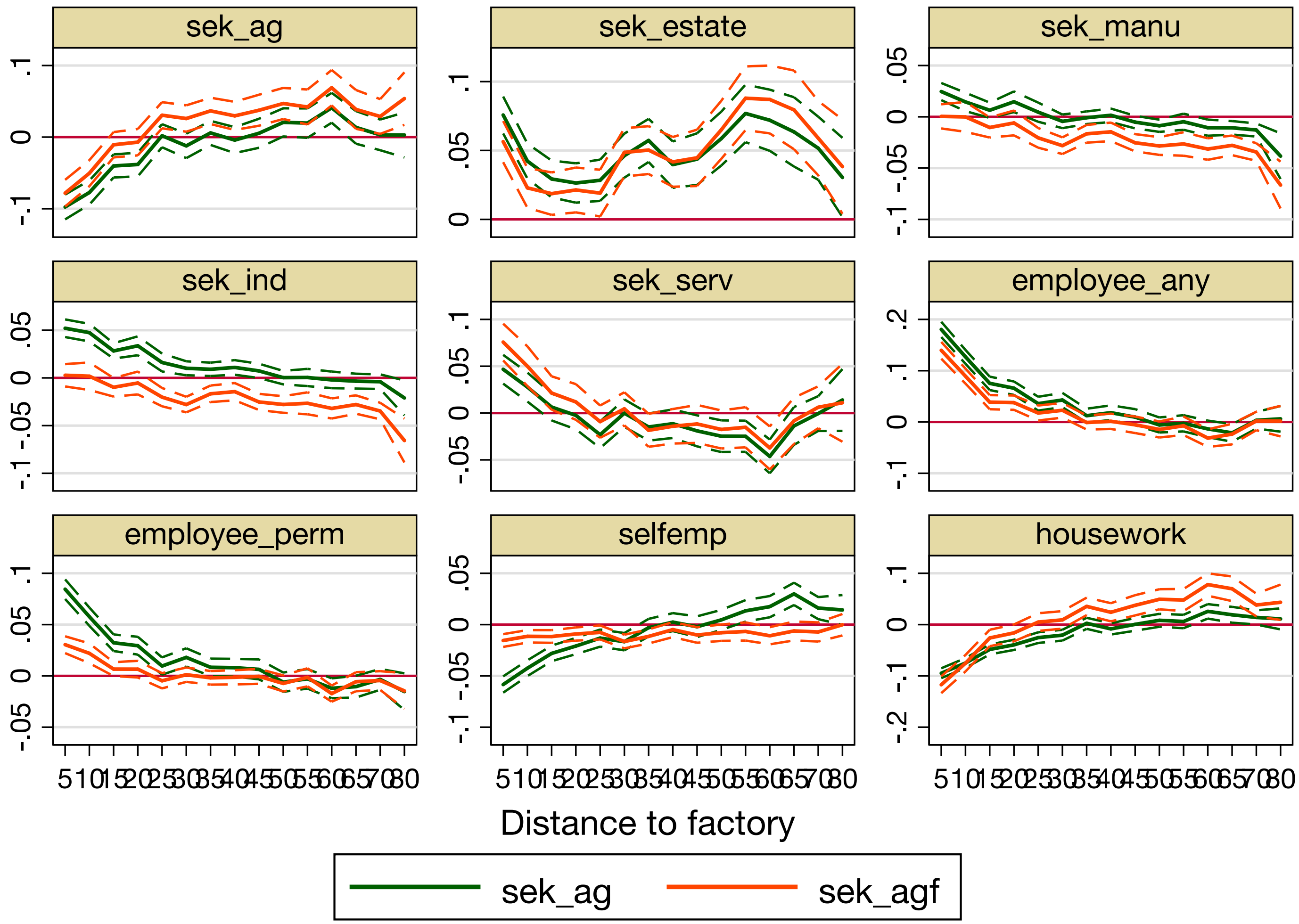
Notes: These figures plot the coefficients from estimating Equation 1 at 5km bins using as a dependent variable different fiscal outcomes reported in PODES 2014. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Confidence intervals are for 95%.

Figure A11: Proximate Adoption and Poverty—No Sample Restrictions



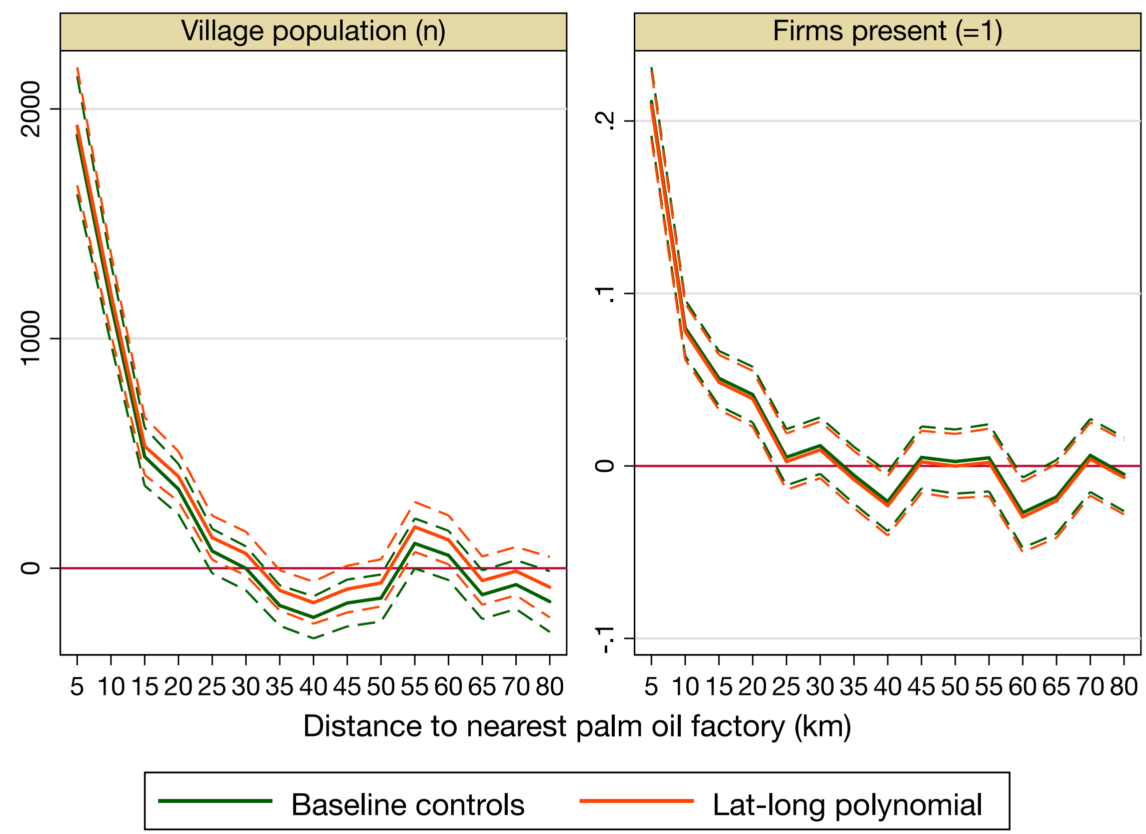
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using the village oil palm acreage and poverty rates measured at the international line as dependent variables. 95% confidence intervals are represented by the dotted lines. Additional controls add to the baseline controls travel time to the nearest city, travel cost to the nearest city, and indicators for river, coast, plains, valleys, forest proximity (in, near, and outside) and forest function. Lat-long polynomial further adds a polynomial in latitude and longitude. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A12: Local Labor Markets—No Sample Restrictions



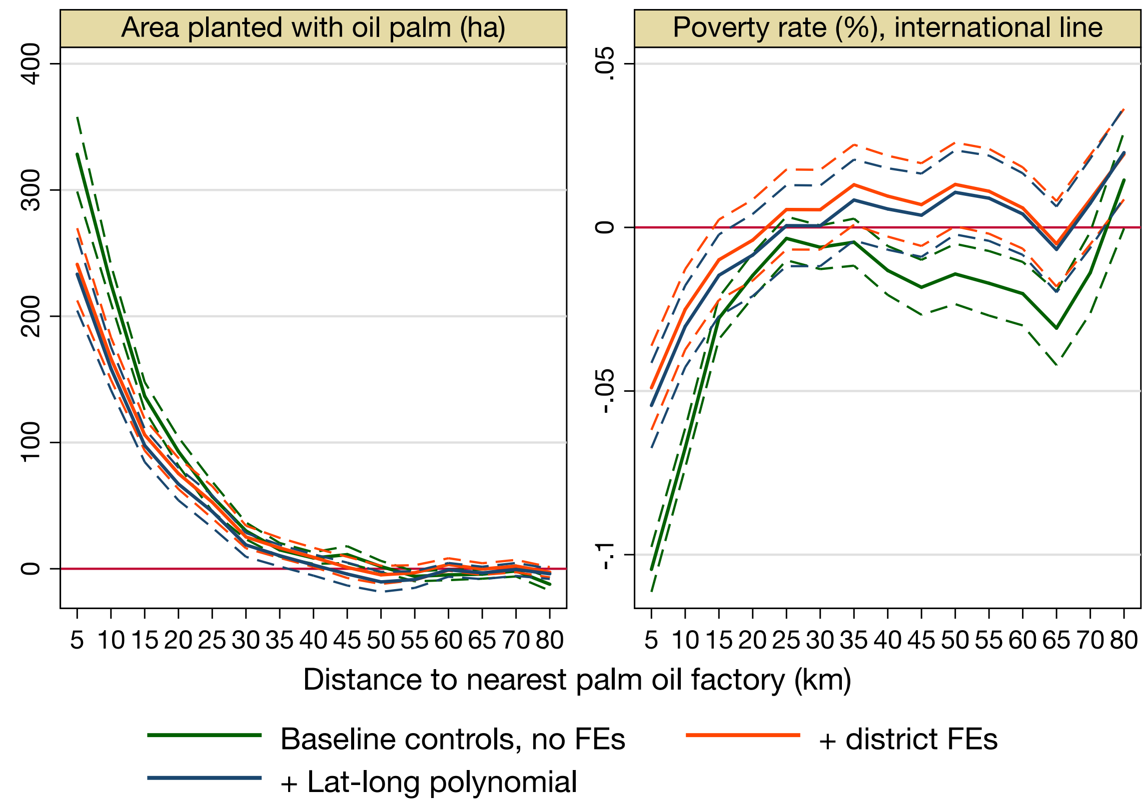
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable the main sector of employment and employment status at the individual level as observed in the 2010 Population Census (10 percent random sample, stratified on village by the author). 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Robust standard errors are clustered on village.

Figure A13: Population and Firms—No Sample Restrictions



Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable village population measured in the 2011 Village Census and the whether a village has a firm present according to the 2016 Economic Census. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

FIGURE A14: PROXIMATE ADOPTION AND POVERTY—NO FACTORY FES

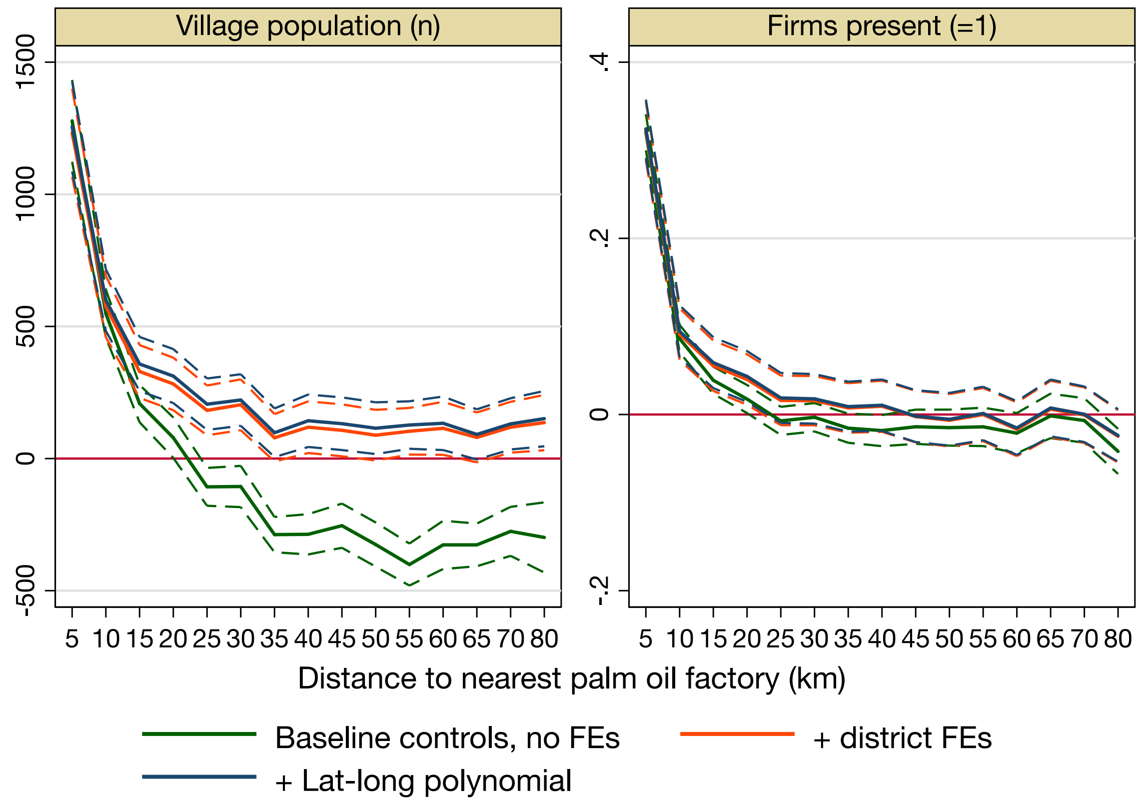


Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using the village oil palm acreage and poverty rates measured at the international line as dependent variables. 95% confidence intervals are represented by the dotted lines. Additional controls add to the baseline controls travel time to the nearest city, travel cost to the nearest city, and indicators for river, coast, plains, valleys, forest proximity (in, near, and outside) and forest function. Lat-long polynomial further adds a polynomial in latitude and longitude. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A15: Local Labor Markets—No Factory FEs

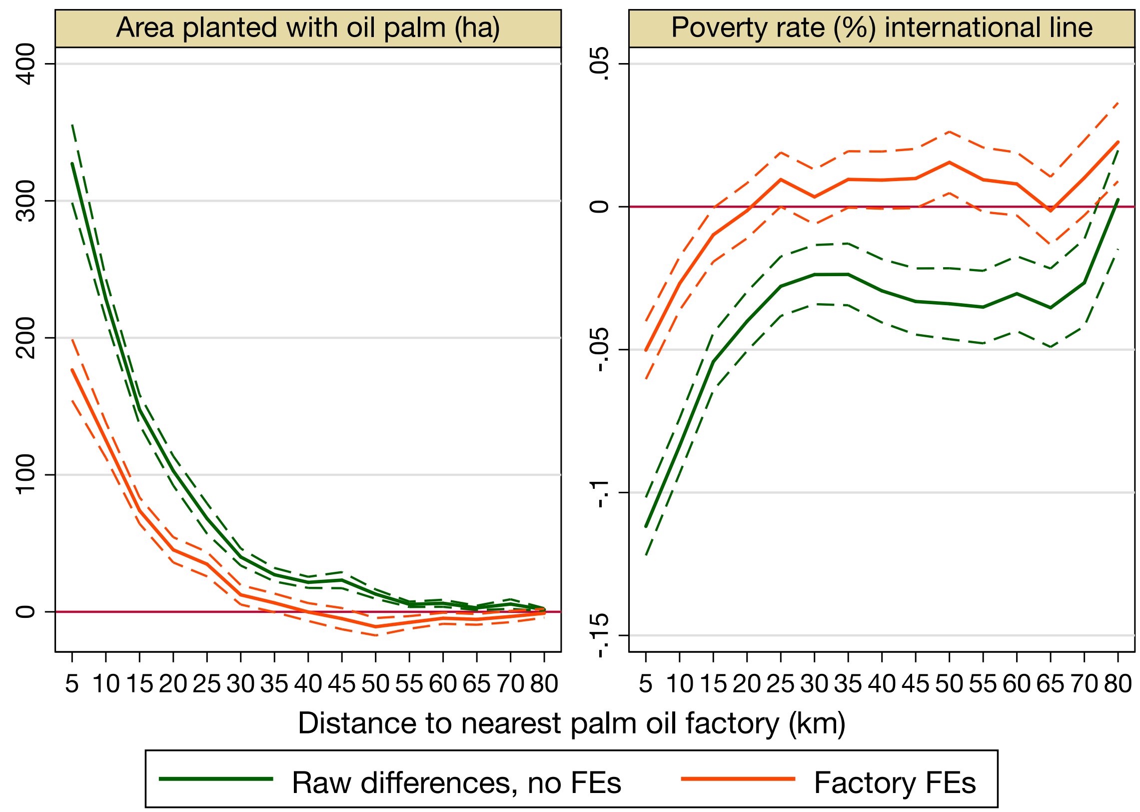
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable the main sector of employment and employment status at the individual level as observed in the 2010 Population Census (10 percent random sample, stratified on village by the author). 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Robust standard errors are clustered on village.

Figure A16: Population and Firms—No Factory FEs



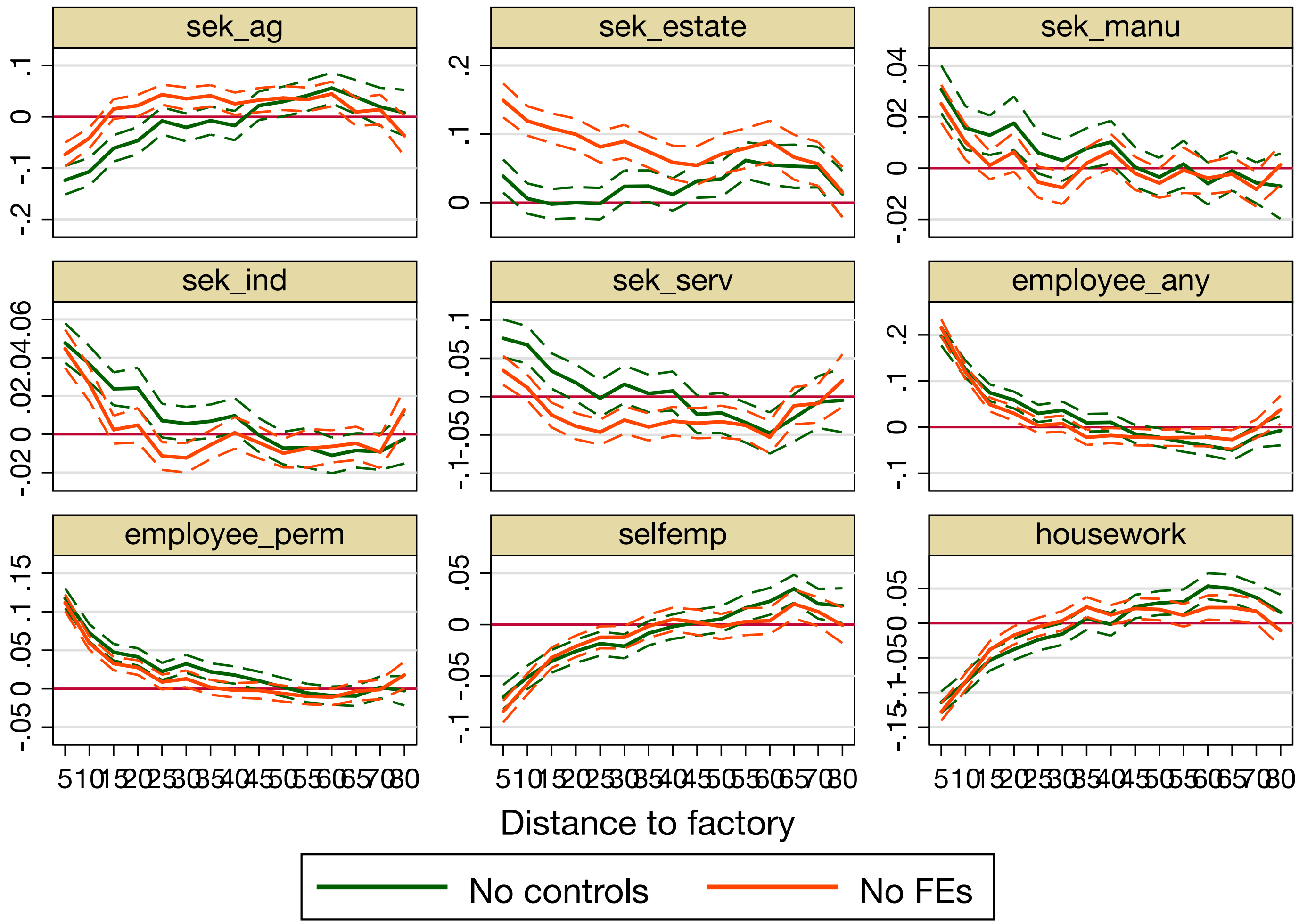
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable village population measured in the 2011 Village Census and the whether a village has a firm present according to the 2016 Economic Census. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A17: Proximate Adoption and Poverty—No Covariates



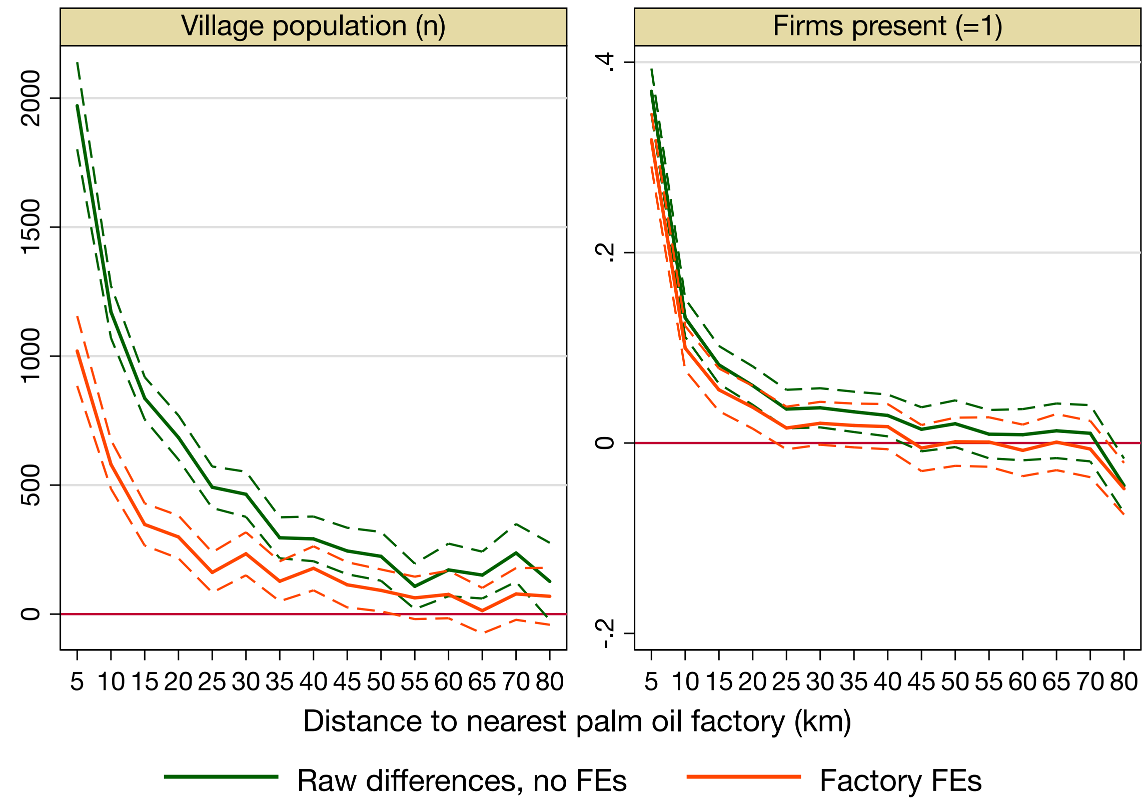
Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using the village oil palm acreage and poverty rates measured at the international line as dependent variables. 95% confidence intervals are represented by the dotted lines. Additional controls add to the baseline controls travel time to the nearest city, travel cost to the nearest city, and indicators for river, coast, plains, valleys, forest proximity (in, near, and outside) and forest function. Lat-long polynomial further adds a polynomial in latitude and longitude. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A18: Local Labor Markets—No Covariates



Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable the main sector of employment and employment status at the individual level as observed in the 2010 Population Census (10 percent random sample, stratified on village by the author). 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java. Robust standard errors are clustered on village.

FIGURE A19: POPULATION AND FIRMS—NO COVARIATES



Notes: The figures plots the coefficients from estimating Equation 1 at 5km bins using as a dependent variable village population measured in the 2011 Village Census and the whether a village has a firm present according to the 2016 Economic Census. 95% confidence intervals are represented by the dotted lines. The sample is all villages within 100 km of a palm oil factory, excluding those that are in major cities (i.e., kota) or on Java.

Figure A20: Kerinici area, Riau—Google Maps Today

Figure A21: Factory Onset—Kayung Agro area, West Kalimantan

(a) 2000, No Factory

(b) 2010, Factory Construction

(c) 2016, Production

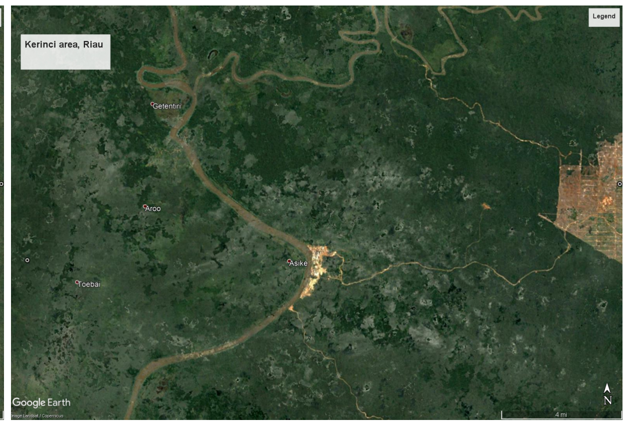
(d) Nearest Road to the East

(e) Road on Google Maps

Figure A22: Aiske area, Papua—LandSat Imagery since 1984

(a) 2000

(b) 2010



(c) 2016

Figure A23: Recent Peri-Urbanization in Bengkalis, Riau—2006-16

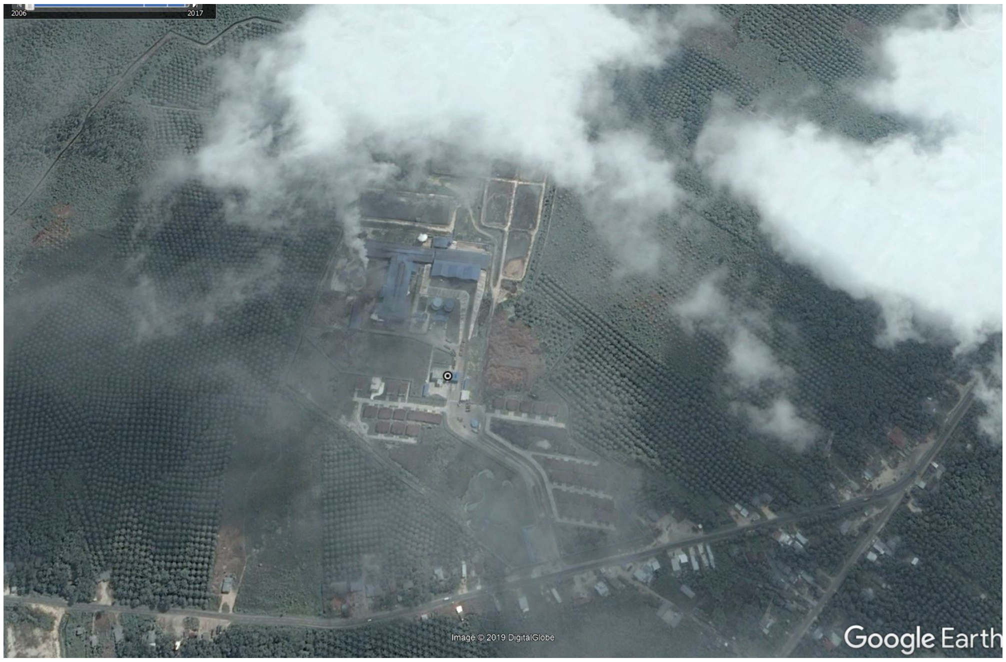
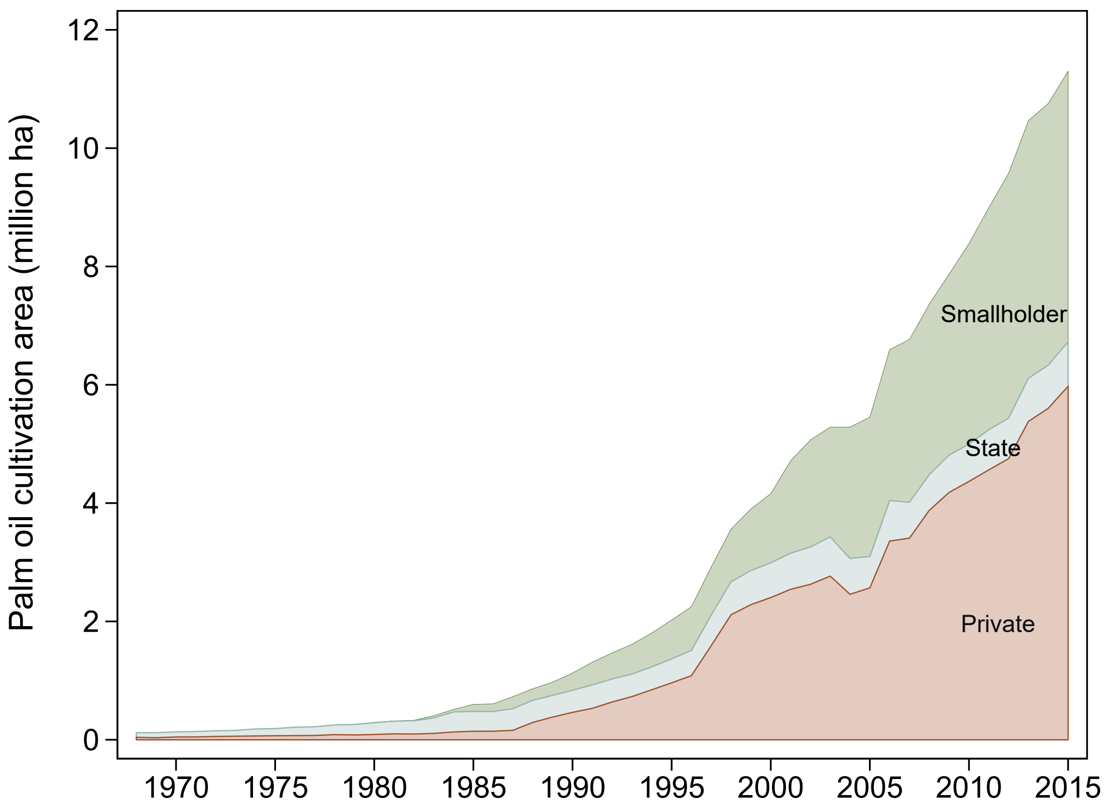
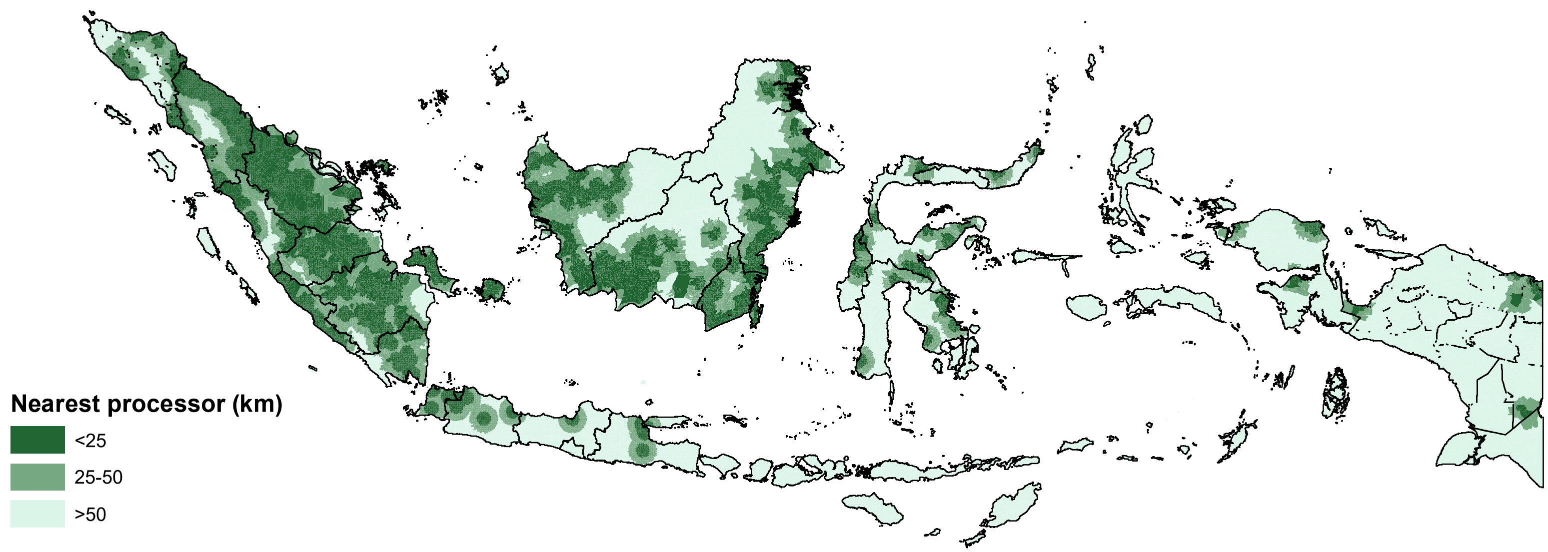


Figure A24: Indonesian Palm Oil Area—1970–2015



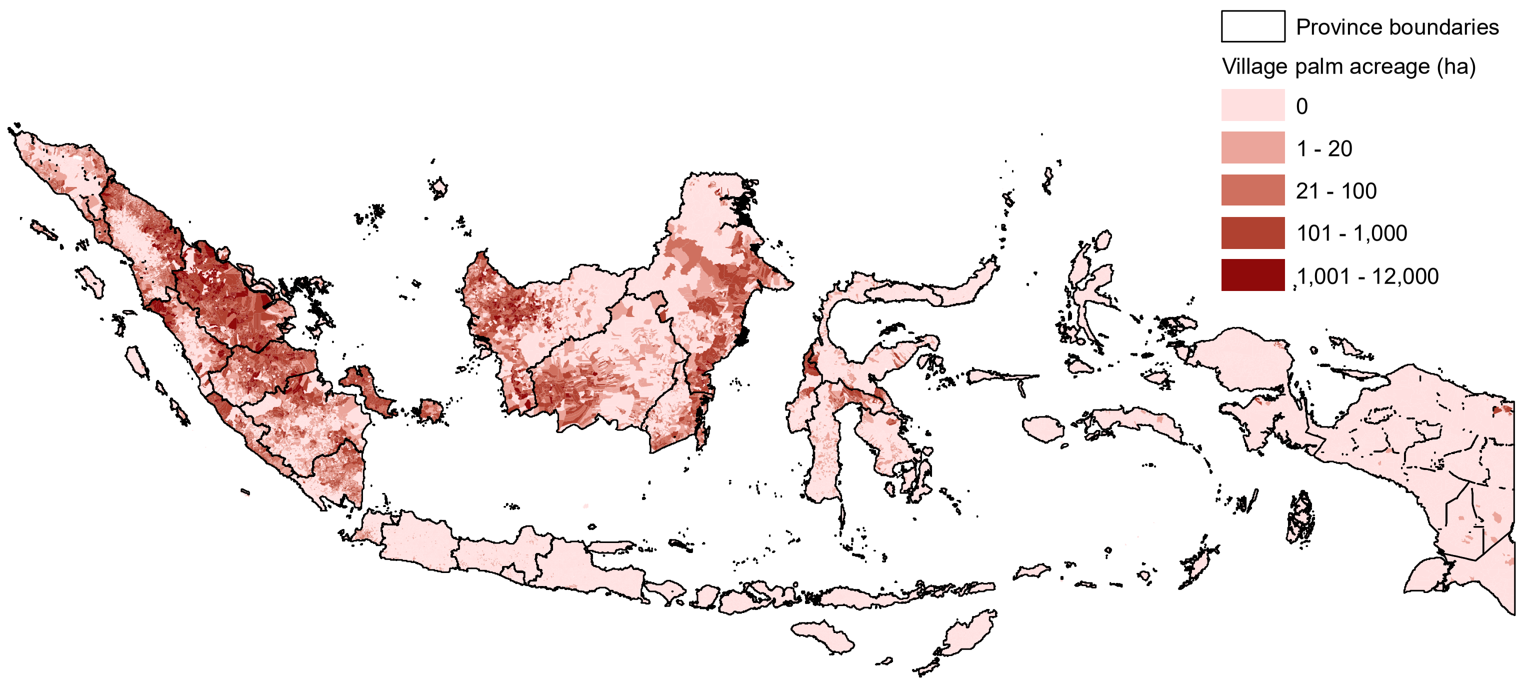
Notes: Data are taken from the Tree Crop Statistics of Indonesia for Oil Palm yearbooks, produced annual by Badan Pusat Statistik (BPS) and the Department of Agriculture of the Government of Indonesia and digitized by the author.

Figure A25: Distance to Nearest Palm Oil Processor From Every Village



Notes: Source: Author's own calculations from the 2016 Economic Census.

Figure A26: Village Oil Palm Acreage



Notes: Calculated from 2013 Agricultural Census.

Figure A27: Processors and Village Oil Palm Acreage—Sumatra

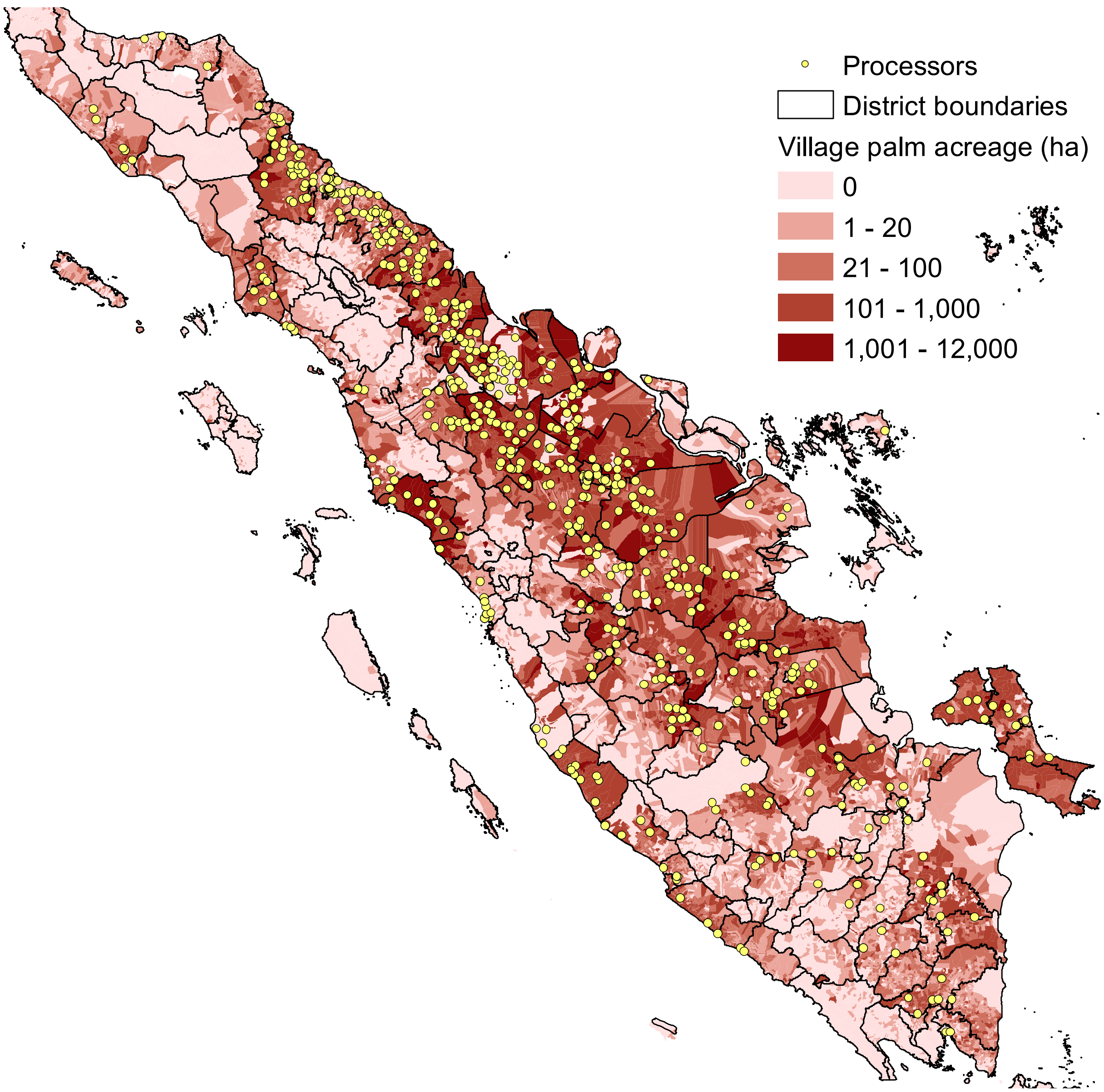


Figure A28: Processors and Village Oil Palm Acreage—Kalimantan

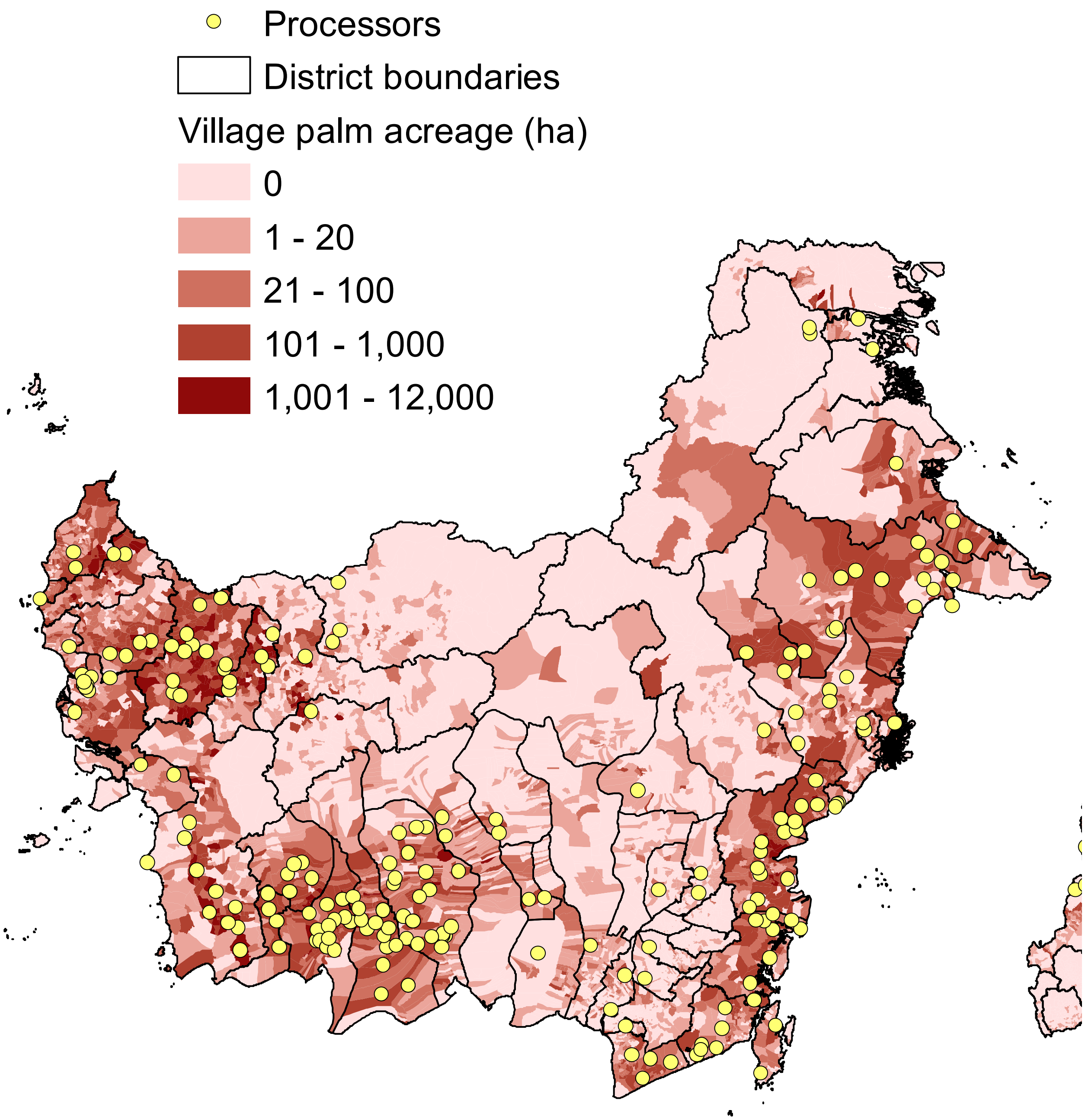


FIGURE A29: PROCESSORS AND VILLAGE OIL PALM ACREAGE—EASTERN INDONESIA