Spillovers to manufacturing plants from multi-million dollar plantations: evidence from the Indonesian palm oil boom*

Sebastian Kraus[†] Robert Heilmayr[‡] Nicolas Koch[†§]

May 21, 2021

Abstract

We estimate spillover effects to local manufacturing plants from the Indonesian palm oil boom, using a stacked difference-in-differences approach. We use new data on the establishment dates and ownership of palm oil mills to identify clean shocks from investments in new plantations. Local plantation booms increased sales and productivity of manufacturing plants, despite increasing blue-collar wages. Using confidential input-output data, we rule out the possibility that this effect is driven by supply chain linkages. Plants increased their share of tradable goods, but produced fewer relationship-specific goods. Local road upgrades point to improved market access as an explanation for this shift.

*kraus@mcc-berlin.net, rheilmayr@ucsb.eu, koch@mcc-berlin.net. We thank Mark Curtis, Ryan Edwards, Sabine Fuss, Kelsey Jack, Krisztina Kis-Katos, Kyle Meng, Sudarno Sumarto, Daniel Suryadarma, Asep Suryahadi, Ping Yowargana, Piotr Śpiewanowski, and seminar participants at MCC Berlin, IRSA in Surakarta, UC Santa Barbara, and at AERE for their helpful comments. We thank staff at the Indonesian statistics agency, BPS, for their trust and excellent support. We are grateful to Jason Jon Benedict, Claudia Günther, Hanif Kusuma Wardani, and Mayang Krisnawardhani for their invaluable research assistance. Sebastian Kraus acknowledges funding by the RESTORE+ project (http://www.restoreplus.org/), part of the International Climate Initiative, supported by the Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU) on the basis of a decision adopted by the German Bundestag.

[†]Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany

[‡]Bren School of Environmental Science & Management, UC Santa Barbara, United States

[§]IZA Institute of Labor Economics, Bonn, Germany

1 Introduction

A positive shift of comparative advantage in labor-intensive agriculture can crowd out industrial growth as workers transition to agriculture. However, large agricultural investments can also lead to positive agglomeration spillovers to unrelated industrial sectors. We use the expansion of palm oil in Indonesia as a quasi-experiment to study the effects of a rapid and large-scale agricultural expansion on an industrializing economy. Palm oil companies typically make greenfield investments of US$100 million to set up a mill and its adjacent oil palm plantations. We show that these investments have, on average, led to positive shifts in sales (15%), labor productivity (13%) and total factor productivity (13%) in local non-palm oil manufacturing plants.

Over the past 20 years, Indonesia has experienced a quadrupling of palm oil production, and an associated dramatic transformation in its rural economy. In 2000, Indonesia still exported more petroleum oil, electronics, garments, and wood products than palm oil. By 2015, palm oil had become Indonesia's largest export with a share of 11% (up from 2% in 2000). In contrast to the Green Revolution, much of this rapid growth in agricultural production was not driven by intensification through technological advances, but rather through extensification, often into natural forests. This extensification was made possible by Indonesia's political and fiscal decentralization process, which accelerated the allocation of land to oil palm and timber concessions (Burgess et al. 2012).

In the absence of dramatic technological change in agricultural practices, labor intensity on Indonesia's industrial plantations has been steady and relatively high. The sector today employs two million people, the large majority of them as plantation workers. Oil palm fresh fruit bunches are primarily harvested and collected manually, and at least one worker is needed for every 6 to 8 ha. Throughout our study period from 2005 to 2015, palm oil has remained a business with a high labor intensity and has exhibited strong complementarity between labor and land.[1] Most labor-saving changes have been achieved through investments in palm oil mill technology, rather than on plantations. Therefore, the arrival of a new palm plantation in a district has put pressure on blue-collar labor markets.

This paper examines how manufacturing plants have reacted to local waves of palm oil expansions in Indonesia. Given that oil palm expansion may simultaneously increase demand for labor and investments in local infrastructure, the palm sector's boom has theoretically ambiguous impacts on other manufacturing

[1]See Appendix D.1 for more detail on innovations and investments in the palm oil sector that impact factor productivity and substitution elasticity.

facilities. We use temporal variation created by the staggered roll-out of palm oil mills to identify spillovers to non-palm oil manufacturing. The establishment of palm oil mills has been at the centre of each wave of local plantation expansions. Palm fruits are perishable and experience significant declines in quality if they are not processed within 24-48 hours. Palm oil mills are typically built for a capacity of 60 tons of fresh-fruit bunches per hour. This means that they need a supply shed of about 10,000 ha. A new palm oil mill and its adjacent plantations, therefore, constitute an investment of around US$100 million.[2] These investments are typically made directly or via proxy by a handful of large palm oil conglomerates. These conglomerates hold a portfolio of concessions often more than double the size of their planted area. We argue that the palm oil conglomerates' decisions on the order in which they use their concessions is plausibly exogenous to local shocks. Their first order concerns are climate, topography, and distance to rivers; they do not rely on local banks, their mills produce their own electricity, and they build their own roads.[3]

Data on the palm oil supply chain is a well-kept secret in Indonesia. Previously, only the location of a subset of palm oil mills has been known. We use a new panel data-set of most palm oil mills in Indonesia, including their establishment dates and ownership structures. Confidential input-output data from the manufacturing census allows us to control for supply chain linkages.

We use a stacked difference-in-differences design, which allows us to examine pre-trends even in our setting with staggered, repeated treatment events. In the research design, we prevent already treated units from acting as controls, while they are still following a different trend. Having control over the comparisons made in the regressions is important. since our treatment does not act as a pure level shifter but exhibits dynamics over several years. We pool all mills that are part of the same wave, i.e., that have been established in the same year, in one treatment group. We call these treatment clusters "cohorts". For each cohort, we restrict our study window to five years before and after treatment. We compare outcomes in the treatment cohorts to controls drawn from manufacturing plants that were not influenced by new palm oil mill investments in the same year. In addition, we exclude plant-year observations from the control group for those years in which we expect control units still to be on a different trend from prior treatment. Thus, the stacked research design allows us to avoid some of the issues that arise from undetected common trends violations and regression

[2]See Appendix D.1 for more details on the investment needed to start a palm oil operation.
[3]See Section 2.2 for more detail on the investment decision-making of palm oil companies

weights on heterogeneous sub-effects in standard two-way fixed effect regressions (Goodman-Bacon 2018).

Our analysis yields four key insights into the impacts of palm oil booms on local manufacturing economies. First, we show that new palm oil mills lead to increases in sales (15%), labor productivity (13%), and total factor productivity (13%) of non-palm oil manufacturing plants, and we rule out the possibility that this effect is driven by upstream and downstream plants. Second, palm oil booms lead to increases in blue-collar wages at non-palm oil manufacturing plants, but migration partially offsets this increase. Third, at the district level, we document growth in tax revenues and increases in the share of asphalt roads. Fourth, using data on all outputs on the plant-level, we show that non-palm oil manufacturing plants increase the share of tradable goods they produce, but decrease the share of relationship-specific goods. This pattern is consistent with improved access to markets due to better transport infrastructure.

We document flat pre-trends, both visually and in regression form. Our results are robust to changing the control group definition, for instance discarding never-treated units and only using variation in treatment timing to identify the effects. We can also check robustness by restricting the estimation sample to the four biggest *private* palm oil conglomerates, which hold concessions for a large portfolio of potential investments, so-called landbanks, and therefore assign treatment more independently from changes in local conditions.

We thus provide a relevant data point to the discussion on the effects of the Indonesian palm oil boom on structural change. The share of manufacturing value added in the Indonesian GDP peaked in the early 2000s and the country has been labeled a case of premature industrialization (The Jakarta Post 2016). Notably, the primary palm oil producing islands of Sumatra and Kalimantan have lagged behind Java in industrial performance, raising the question of whether the growth of the palm oil sector has crowded out other industrial activity. We investigate one local channel of such a potential resource curse and find the contrary: the average incumbent non-palm oil manufacturing plant experiences positive spillovers from plantations. Our counterfactual cannot teach us about the industrialization path Indonesia as a whole would have taken without the palm oil boom, but it casts doubt on the hypothesis of negative local agglomeration externalities as a driver of an industrial slowdown at the aggregate level.

Related Literature Our results contribute to three research domains in resource, development, and agricultural economics.

First, we add to the literature on resource curses by investigating whether Indonesia's oil palm boom has contributed to the country's slow-down in local structural change. The palm oil industry relies on a decentralized network of mills because palm fruit needs to be processed within a day. Thus, incumbent firms from other sectors in the surrounding areas are exposed to new processing facilities and their potential agglomeration effects similar to the effects found by Greenstone et al. (2010) of new "million dollar" manufacturing plants on incumbent plants in the US. However, local up- and downstream supply chain linkages in the palm oil industry are limited and, during our study period, 70 to 85% of crude palm oil was exported. In this regard, our empirical context is similar to that of point resources, such as oil wells or mines.[4] Allcott and Keniston (2018) use the US census of manufacturers to show that manufacturing is not crowded out by local wage increases during natural resource booms. Cust et al. (2019) study oil and gas windfalls in Indonesia and show that manufacturing plants resist the Dutch disease effect of higher wages; on average, they manage to increase productivity and even output. While the prior literature focuses on less labor-intensive resource contexts, we study a natural resource-based boom with a very high labor intensity. Our finding that local industrial performance increases, despite the much higher labor demand created by palm plantations compared to oil and gas wells, casts doubt on local structural change as a driver of a resource curse, even in labor intensive resource use contexts.

Second, we build on a growing literature examining the socioeconomic effects of the Indonesian palm oil boom. Oil palm plantations have lifted 1.3 million people out of poverty (Edwards 2019a), are linked to a higher density of small businesses (within 20 kilometers of the mills) (Edwards 2019b), are associated with decreases in household fertility (Kubitza and Gehrke 2018), increases in household consumption, calorie consumption, and dietary quality (Euler et al. 2017), but are also linked to lower formal employment and wages (Coxhead and Shrestha 2016). Prior research on the agglomeration effects of palm oil was based on a cross-section of palm oil mill locations (Edwards 2019b). We use panel data that includes mill ownership and establishment dates and the Indonesian manufacturing census to investigate the dynamic adjustment of non-palm oil manufacturing plants in reaction to new palm oil investments. We know which large corporate groups are behind the investments and can, therefore, investigate treatments that are plausibly exogenous to local economic conditions. We use

[4]For an overview of the literature on the local economic impacts of resource extraction see Cust and Poelhekke (2015)

plant-level input-output data to remove direct effects through supply chains. In addition to confirming and strengthening prior results that oil palm mills have improved local economic conditions, our analysis provides new insights into the palm oil boom's important, structural impacts on local manufacturing output, productivity, wages and tax revenues.

Third, we evaluate the local structural change effects of an expansion of the supply of fertile land for agriculture. The Green Revolution in India (Foster and Rosenzweig 2004, 2007; Moscona 2019), colonial sugar factories in Indonesia (Dell and Olken 2020), and the soy boom in Brazil (Bustos et al. 2016) are other, well-studied cases of the links between agricultural expansion and industrial growth. In Brazil, new soy farms have also benefited from cheap land at the deforestation frontier. However, in contrast to the Indonesian oil palm sector, employment in the Brazilian soy sector has halved due to the introduction of new technologies (GMOs and automation). Bustos et al. (2016) have shown that the induced release of labor has led to *locally* increased employment in the manufacturing sector. The Indonesian palm oil boom was not accompanied by comparable technological advances. In palm oil districts we find an overall shift towards employment in agriculture. However, associated increases in blue-collar wages need not lead to a contraction of the manufacturing sector – we find increases in sales from local manufacturing plants that compete for blue-collar workers with the new palm oil mills and plantations. We attribute the sales effect to agglomeration spillovers resulting from improvements in infrastructure and increases in local tax revenues. These spillovers are reflected in plant-level productivity increases that are similar in magnitude to the increases in sales.

2 Empirical strategy

We estimate the effect of the establishment of a palm oil mill on incumbent manufacturing plants from non-palm oil related industries.

Our main empirical challenge is the endogenous placement of palm oil mills. Our research design leverages variation in the timing of palm oil mill establishments. We compare manufacturing plants that are exposed to a new palm oil mill to manufacturing plants in areas without new investments around the same time. For this to be a credible counterfactual in the presence of a set of fixed effects and parallel pre-trends, we need to provide supporting evidence that there are no local shocks that coincide with mill adoption and that would have put manufacturing plants on a different trend in the absence of the treatment.

Our identifying intuition here is that palm plants are a part of large conglomerates that decide independently from local shocks when to make use of the concessions they hold in their land banks (see Appendix D.3 for a discussion of the concession licensing process). They set up their plantations in places where climate, topography, and distance to rivers are suitable. They are independent from local funding, build their own roads, and their mills typically generate their own electricity. Since the palm oil sector is concentrated, but ownership structures are informal and opaque, our main robustness checks focus on samples of plants that are officially known to be part of large palm oil firms.

2.1 Main specifications

We use a stacked difference-in-differences design to estimate both point estimates and leads and lags of spillovers from palm oil shocks on local incumbent manufacturing plants that are not part of the palm oil business and supply chain.

Since palm oil shocks happen in a staggered manner, a standard two-way fixed effects regression comparing changes in pre- and post-treatment outcomes between different subsets of treatment and control group would be biased, if treatment effects are dynamic (Goodman-Bacon 2018). Intuitively, this would happen because treatment puts manufacturing plants on a different trend, rather than only shifting its levels (see Appendix 6 for a more comprehensive discussion). This leads to violations of common trends if they act as a control group for manufacturing plants treated later.[5] Since we expect at least some dynamics in the adaptation of non-palm manufacturing plants to their new business environment, for example due to investment planning and hiring lags, we cannot rule out that part of the effect builds up over time.

Instead of estimating a standard two-way fixed effects model regressing manufacturing plant outcomes on a running total of palm oil mills in each district or the corresponding total palm plantation area, we therefore use the establishment of a new palm oil mill as a treatment event. Specifically, we identify our treatment effect by comparing manufacturing plants whose treatment status changed in the respective year, which de Chaisemartin and D'HaultfŒuille (2018) call "switchers", to manufacturing plants who were not exposed to changes in nearby palm oil infrastructure in an exclusion window around this event.

We illustrate the construction of our stacked sample in Figure 1. Each year

[5] Athey and Imbens (2018) call the assumption of no dynamic treatment effect "invariance to history", i.e., potential outcomes are only influenced by the fact that a unit is treated and not by how long it is treated.

within our study period (2005-2015) defines a cohort of palm oil mills (see first row in Figure 1). We create individual data sets for each of the cohorts of palm oil mills. We restrict these sub-samples to observations from five years before and five years after the cohort's year, since this is the event study window for which we estimate leads and lags (see second row). We stack the cohort sub-samples in event-time for a pooled regression. Manufacturing plants are assigned to a cohort's treatment group if a palm oil mill has been established in their district in the respective year. They are eligible for the control group if there was no new palm oil mill in their district in the cohort's year.

[Insert Figure 1 here.]

The stacked design also allows us to define an observation exclusion window W_{ex} for observations to obtain cleaner shocks and controls (see third row). We ensure that control units are not influenced by past treatment by excluding manufacturing plant-year observations if there is another treatment event occurring anytime in the preceding three years and in the observation-year itself. We also run robustness checks extending this post-treatment exclusion window to six years and introducing a pre-treatment exclusion window of three years. The latter ensures that observations are free of major anticipation effects which can appear, since in order to ensure a steady supply of fresh fruit bunches for an operational mill, palm oil companies often begin to clear land up to six and plant palms four years prior to mill establishment. Thus, anticipation effects, if relevant, should typically appear around three years before the establishment year of mills.

Note that manufacturing plant-year observations appear in several of these cohort sub-samples. They can be in the treatment groups of several cohorts, since the establishment of a palm oil mill is a recurring event in any district. They will also appear in the control groups of other cohorts, if they have a large enough break from palm oil mill establishment to be considered controls that do not experience any (dynamic) treatment effects in the event window.

Since palm oil mill establishment is a recurring event, there are trade-offs between the balance of cohorts' treatment and control groups, the length of the event window, and the associated necessary exclusion of units that can be on a different trend because of previous treatment or anticipation. For instance, if we choose a wide event window and only include observations after a break from treatment of the same length, we tend to select booming districts out of the sample and therefore estimate results on a sample of younger and more mature palm areas. If manufacturing plants from the latter areas are on systematically different trends, or even differently exposed to local shocks endogenous to treatment timing, we may increase omitted variable bias. In robustness checks, we also show results for a larger event window of eight years.

Note that we cannot estimate a standard, "non-stacked" event study on a one-off, staggered treatment, since manufacturing plants can be exposed to several shocks over the study period and the scarcity of data on palm oil mills does not allow us to go back in time to the first establishment of a palm oil mill in a sufficient number of districts.

The stacked difference-in-differences approach has previously also been used by Cengiz et al. (2019) and Deshpande and Li (2019).[6] We estimate this OLS regression at the plant-year level:

$$\begin{aligned} \ln Y_{\text{idrycs}} = \lambda_i + \mu_{ry} + \nu_{sy} + \delta_0 \text{Treated}_{dc} \times C_c + \\ \Sigma_\tau \delta_1 D^\tau_{cy} \times C_c + \Sigma_\tau \delta_\tau \left(\text{Treated}_{dc} \times D^\tau_{cy} \right) + \epsilon_{\text{idrycs}}, \quad (1) \end{aligned}$$

where Y_{idrycs} is an outcome of interest (sales, labor productivity, TFP, wages, product portfolio variables) for plant i, in sector s, in district d, in island r, in year y, in cohort c. Our palm oil mill cohorts go from 2005 to 2015.

Treated is a dummy that indicates whether a manufacturing plant's district is treated in the treatment year of a cohort. We estimate 4 leads and 5 lags around treatment. The D^τ_{cy} are binary indicators that are 1 if year y is τ years before or after the treatment year of the cohort in which the observation appears. C_c indicates whether an observation is part of a cohort. In all these event-study type specifications, the reference year is the year just before the establishment of a mill, when τ equals -1.

λ_i are manufacturing plant fixed effects, ν_{sy} are five-digit industry-year fixed effects that capture unobservable changes common to manufacturing plants that have the same main product.[7] μ_{ry} capture time-varying unobservables at the island level. The Indonesian islands are naturally separated economies with their own electricity grid, port infrastructure and political dynamics. The main Indonesian islands where palm oil is grown are Sumatra, Kalimantan, Sulawesi, and Papua.[8]

The parameter of interest is δ_τ. It captures the difference in outcomes over

[6]See Appendix for a more extensive discussion and additional applied examples of this design.

[7]We also run specifications with fixed effects at the island-industry-year level (see Table 3)

[8]The main island, Java, with the Indonesian capital region around Jakarta, does not play an important role in our analysis, since it mainly hosts refining and logistics infrastructure, but only two palm oil mills.

time between manufacturing plants in the same industry and on the same island, with the only difference being that some are located in districts with a new palm oil mill in a given year, while others do not experience such a palm oil shock in the exclusion window of three years before the treatment year.[9] We show this parameter over time in regression form and charts in the following sections.

Compared to an event study, the use of a control group enables us to remove time trends that are common to manufacturing plants in event-time in addition to standard calendar-time fixed effects. When we create interactions of the cohort indicator C_c with the Treated_{dc} indicator and the event-time indicators D^{τ}_{cy} respectively, we use the same fixed effects that we would be using in individual difference-in-differences for each cohort, thereby effectively estimating effects within cohorts.

The interaction of D^{τ}_{cy} and C_c removes cohort-specific unobservables that appear in event-time, rather than calendar time. When we just include D^{τ}_{cy}, we only remove this variation over the pooled and stacked sample of pre- and post-time steps around events. The interaction of Treated_{dc} and C_c removes time-invariant differences between treatment and control groups of each cohort. This includes time-invariant unobservables that could be driving outcomes and selection into earlier or later treatment.[10] When we just include Treated_{dc}, we control for these differences only between the pooled treatment group and the pooled control group.

We also run the following pre-post specification to capture the average treatment effect over the five years after a mill is established compared to the preceding four years (leaving out the year before treatment):

$$\begin{aligned}\ln Y_{\text{idrycs}} = {} & \lambda_i + \mu_{ry} + \nu_{ys} + \beta_0 \text{Treated}_{dc} \times C_c + \\ & \Sigma_{\tau} \beta_1 D^{\tau}_{cy} \times C_c + \beta \left(\text{Treated}_{dc} \times \text{Post}_{cy}\right) + \kappa \left(\text{Treated}_{dc} \times \text{Zero}_{cy}\right) + \epsilon_{\text{idrycs}} \quad (2)\end{aligned}$$

We dummy out the year of the mill establishment (using the interaction of Treated_{dc} and $Zero_{cy}$), since we only know the year, when a mill appears in the official records, but not the exact timing. The coefficient of interest is β. It captures the difference-in-differences estimate averaged over the five years before and after treatment.

[9] We also show results with a more sparse set of fixed effects in Figure 5

[10] See Appendix 6 for a detailed discussion of treatment effect heterogeneity with regards to regression weights in dose-response two-way fixed effects specifications.

2.2 Identifying assumptions

Our model yields a causal estimate of the spillover effects of new palm oil mill establishment contingent upon four main assumptions.

Parallel trends Our core identifying assumption is that, within the sub-populations created by our fixed effect structure, the manufacturing plants in districts with a new palm oil mill would have seen the same sales and productivity growth as those manufacturing plants located in districts where no new mills were established in that year, excluding their observations that we expect to be on a different trend because of prior treatment. Since we model our manufacturing plant outcomes (sales, total factor productivity, labor productivity, wages) as logs, we assume that outcomes of treated and untreated manufacturing plants would have grown at the same rate, rather than in absolute terms. Since manufacturing plants are heterogeneous in their baseline sizes and productivity, this is the more plausible parallel trends assumption.

No anticipation For our difference-in-differences estimate, based on the difference between the means of the five years before and after treatment, to be unbiased, we need to assume that there are no anticipation effects, since this would change the trajectories of both our treatment and our control groups. Intuitively, this should attenuate our effect, except if there is an Ashenfelter-type dip in outcomes pre-treatment. Our stacked design allows us to check the robustness of our results to excluding observations suffering from anticipation from our study sample by modifying both the event window and the exclusion window.

No endogenous timing We also need to assume that treatment timing is not endogenous. Our estimates would be biased if there was reverse causality or a third factor driving both treatment and outcome. Therefore, we need to argue that, conditional on fixed effects, adoption dates are not driven by any omitted variable that also drives manufacturing plant outcomes. We also have to rebut the argument that the performance of manufacturing plants drives the adoption date of palm oil mills; for instance, through a lending channel.

Our identification strategy builds on the fact that most palm oil mills are part of large corporate groups that own so-called land banks with a portfolio of potential mill locations. Only in mature palm oil markets, such as Riau, a third of all mills operate independently from large concessions as stand-alone mills, sourcing from independent smallholders (Jelsma et al. 2017). Palm oil mills are established preferentially where palms grow optimally, where land is less hilly and where distances to rivers are shortest.

There are five features of the palm oil sector that make us confident that decisions in the corporate headquarters, as to in which cohort a palm oil mill is placed, are exogenous to changes in local conditions.

First, the decision of where to locate a palm oil mill is mainly based on time-invariant factors, such as climate, soil, topography and distance to rivers. Edwards (2019b) shows a link between attainable oil palm yields and plantation share in districts. He also cites industry actors who describe land suitability as their first-order concern in investment decisions.[11]

Second, the timing of palm oil mill construction is comparatively sheltered from local political economy dynamics. Since the country's political and fiscal decentralization in the beginning of the 2000s, districts have held wide-ranging powers over land allocation (Burgess et al. 2012). Therefore, there could be political economy forces at work that are difficult to measure and that could be driving both land allocation and the performance of non-palm related manufacturing plants, even after removing fixed effects. This form of omitted variable bias could be relevant for palm plants' investment in concessions, but much less for the precise point in time when a mill begins to operate.[12]

Palm oil companies typically acquire land and hold it as an option to build plantations and mills. This is also one of the key assets that they advertise to investors. For instance, the firm Golden Agri-Resources holds 690,000 ha of unplanted land, which is more than the total size of its existing plantations. Palm oil is an economically concentrated sector. A few companies, such as Golden Agri-Resources (Sinar Mas), Salim Group, Wilmar, Sime Darby, and Astra Agro Lestari, own most large plantations in Indonesia, either directly or through opaque shareholder or financing structures.[13] Within these large firms, investment decisions are made within a large portfolio of mill options. Therefore, they are less likely to be directly linked to local political economy shocks or to exhibit a uniform lagged pattern with the attribution of a concession. Our data on corporate ownership of

[11]The land and climate suitability-driven placement of palm oil investment is similar to other sectors whose location decisions are based on factors largely orthogonal to other sectors, such as natural amenities like nearby islands, white sands, or archaeological ruins for tourism (Faber and Gaubert 2019).

[12]In Online Appendix D.3, we discuss that, even in the case when a palm oil company in our sample starts developing its plantation immediately after getting a concession, the timing of the start of operation is unlikely to be driven by local factors that could also be driving the performance of unrelated manufacturing plants, because of idiosyncratic delays in the concession licensing process.

[13]According to an industry insider, the Indonesian Ministry of Agriculture can only "ascertain the ownership of 30 percent of the private companies in the sector" (Baudoin et al. 2019).

palm oil mills allows us to run robustness checks on a sample of larger groups of more than five mills, among which many already own mills on both Sumatra and Kalimantan. We can also restrict the sample to the largest groups, which own more than 25 mills in a regionally diversified portfolio (see Table E.1 for a breakdown of the number of mills owned by large palm oil companies).

Third, most palm oil mills and plantations, except for independent smallholder mills, have been independent from local demand in the study period, since palm oil has largely been for export. Our sample consists of manufacturing plants that are neither part of the palm oil industry, nor directly upstream or downstream to a palm oil mill with a workforce larger than 20, that typically produce for the national and global market. Only a few are businesses further upstream or downstream of palm oil businesses, so adoption decisions are unlikely to be driven by the local evolution of the palm oil sector. During the study period, Indonesia expanded its downstream part of the palm oil supply chain and built refineries, which could be driving local demand in our study. However, refineries are typically built only in industrial centers, typically close to ports, which is captured by region-year fixed effects. Our robustness checks with samples excluding non-palm oil areas, and therefore urban industrial centers, also help counter this concern.

Fourth, relevant infrastructure investments are either made specifically for plantation projects, since they are typically in remote areas, or happen on higher geographical levels that are captured by region-time fixed effects. Palm oil mills have to be located close to plantations, since fresh fruit bunches of oil palms have to be processed within 24 hours. Therefore, mills are found in remote parts of each district. For instance, Gatto et al. (2015) observe that palm plantation area increases with distance from all-season roads. Palm plants typically build their own road networks[14] and mills can produce their own electricity with generators and out of residues[15]. Therefore, they are unlikely to base their decision to build a mill on highly local infrastructure investments by other manufacturing plants or by the government. If government investments coincide with palm oil investments, they are either on a higher geographical level that is covered by fixed effects or can

[14]Palm plantations are set up in 100 ha blocks with collection roads, sometimes even including new bridges, for the trucks transporting fresh fruit bunches at 250 m intervals (Corley and Tinker 2016). Local governments are not involved in developing feeder roads to smallholder plantations.(Jelsma et al. 2017)

[15]The manufacturing census includes information on electricity generation. 80 % of crude palm oil producing plants (mills and refineries) have their own generators, 60 % do not buy any electricity from the grid. Among the remainder, most plants can be considered refineries, which are typically located in more industrialized areas, rather than mills, but the manufacturing census data does not allow clear differentiation between them.

typically be considered a result, rather than a cause, of a local palm oil expansion.

Fifth, mills and their large initial plantations are not financed through local banks. While a palm oil mill in itself is a comparatively simple facility, a plantation area of 10,000 ha has to be set up to run a profitable mill. Including the mill, a hectare of plantation costs US$10,000. Therefore, the total upfront investment is around US$100 million (Byerlee et al. 2016). This type of finance can only be mobilized by large corporations, which are often listed on stock exchanges and collaborate with supra-regional banks. Also, most of the banks that have a network in rural areas are state-owned and operate at the national or at least regional level. Therefore, positive local shocks should be smoothed out of their lending activities.

If there is no omitted variable driving adoption dates and manufacturing plant outcomes, parallel pre-trends are a clear indication that parallel trends hold for potential outcomes after treatment, too.

SUTVA: Limited spillovers between districts We base our estimates on comparisons of manufacturing plants on the same island; they would be biased if there were substantial spillovers between treated and control areas. We run our analysis at the district level, districts being local economies and commuting zones in Indonesia. Nevertheless, there might be spillovers between districts because of labor migration, plant relationships or shifting government priorities. We use the following strategies to guard against these types of spillovers driving our result:

First, in Indonesia, some cities are their own districts: We expect them to experience spillovers from neighboring districts' palm shocks. We run robustness checks merging cities with their surrounding rural districts, excluding those cities that are on the border between districts from the sample. We also show robustness checks using only never-treated plants as controls.

Second, we run our main specifications with region-time fixed effects on the island level, in order to keep the pool of control units reasonably large to limit attenuation bias from treated manufacturing plants that are located in control districts.

Third, we also check for spillovers by running our baseline specification for neighboring districts of treated districts, only excluding the treated districts from the sample.

3 Data

We combine a new panel dataset of the location and the establishment date of palm oil mills in Indonesia with the Indonesian manufacturing census. Our treatment is the establishment of a palm oil mill in a district between 2005 and 2015. Our main outcomes are manufacturing plant-level sales, labor productivity, and total factor productivity. We examine margins of adjustment with plant-level data on all outputs and use data on all inputs to clean data necessary for TFP estimation. In further analyses, we also use district-level data on population size, employment, roads, public investments in infrastructure, and oil, gas and mining revenues.[16]

Palm oil mill panel Our treatment variable indicates whether any palm oil mills were established in a given district in a given year. Palm oil mills are a critical piece of the palm oil production system, and serve as a focal point for oil palm plantations. However, data on the existence, location, licensed capacity, ownership and establishment date of palm oil mills is maintained at the level of individual provinces and, as a result, official, exhaustive data describing these facilities is not publicly available. To fill this gap, we build upon a recently released database detailing the locations of 1150 palm oil mills, representing nearly the entirety of the sector.[17] We supplement this database with data collected and digitized from provincial offices of Indonesia's plantation agency (*Dinas Perkebunan*) as well as corporate reports (Heilmayr et al. 2020), to add attributes detailing the date that a mill was established, the parent company and the corporate group with majority ownership over each mill.

[Insert Figure 2 here.]

In aggregate, our expanded dataset incorporates administrative records on the establishment dates of 533 of the 1150 palm oil mills. 381 of these mills were established between 2005 and 2015, which is our study period. We note that accurate establishment dates prior to our study period are important to minimize common trends violations by removing observations that fall into the exclusion

[16]See Appendix C for additional detail on the construction of the individual variables.

[17]The base sample of our panel is a merge of existing mill location data from researchers and NGOs, the "universal mill list" (World Resources Institute et al. 2019). Data comes from World Resources Institute (WRI), Rainforest Alliance, Proforest, Daemeter, Trase, Earthworm Foundation, Auriga, CIFOR, Transitions, UC Santa Barbara, and the University of Hawai'i. A published version of this merged base sample can be accessed at: http://data.globalforestwatch.org/datasets/universal-mill-list.

window of three years before a cohort's treatment year.[18] For robustness checks, we collect 368 additional establishment dates from secondary sources, such as company reports and satellite imagery.[19] Many of these sources allow us to assign a date range, rather than a precise measure of the mill's establishment date. Therefore their inclusion creates additional statistical noise, but helps us rule out that treatment effects are driven by the fact that units are on different trends from previous treatments. We were unable to determine the establishment date for the remaining 249 mills.

[Insert Table 1 here.]

Figure 2 shows the spatial and temporal distribution of mill investments over districts in Indonesia. Many of our establishment events happen in the same districts, since only 128 Indonesian districts have oil palm plantations. Most mills are on the islands of Kalimantan (the Indonesian part of Borneo) and Sumatra. Only two mills are on the main island, Java. Sumatra (the island in the West) has a higher share of old palm oil mills than Kalimantan (the island in the North). As shown in Table 1, each palm oil mill cohort from 2005 to 2015 contains between 9 and 18 treated districts and between 259 and 277 control districts. The cohorts with the largest treatment groups between 2011 and 2014 correspond to the peak in palm oil mill growth in the full sample, before cleaning and stacking shown in Figure 3. Most large groups have split their investments between Sumatra and Kalimantan. Even among smaller groups with more than five mills, many have investments on both islands. In Table E.1, we show a breakdown of the number of mills of corporate groups by islands.

[Insert Figure 3 here.]

Manufacturing census Our main outcome variables (sales, labor productivity, and total factor productivity) measure the performance of manufacturing plants that are not part of the immediate palm oil supply chain. We take these variables from the Indonesian manufacturing census (IBS or in the economics literature also Statistik Industri), which is collected annually by the national statistics agency (Badan Pusat Statistik (BPS) 2018). We also obtained confidential manufacturing plant-level records of all inputs used in production, all outputs sold and their

[18]For robustness checks that remove potential anticipation effects by setting the exclusion window to three years after the cohort's treatment year, we also use palm oil mill establishment dates beyond our study period up until 2018.

[19]See further explanations on the robustness check samples in Section C.2 and Figures 5, F.1, and F.2 for the corresponding coefficients.

destination country. This information is available both in physical and in monetary terms. We use this previously unavailable data to investigate the margins of adaptation of manufacturing plants, i.e., the share of tradable goods and the share of relationship-specific goods a plant produces.

The Indonesian manufacturing census includes all manufacturing plants above 20 workers, amounting to, on average, around 20,000 plants during the study period.[20] Besides industrial manufacturing, it includes agricultural processing and manufacturing services. We remove all palm oil mills, refineries, and other directly connected parts of the palm oil supply chain from our sample to investigate spillovers to non-palm oil manufacturing plants. In particular, we exclude all plants that produce crude and refined palm oil themselves. Crude palm oil is the largest single product in our raw data when adding up sales at the nine-digit commodity level. In contrast, there are only a few upstream and downstream plants of palm oil mills in the raw data, e.g., no plants that list fertilizer as one of their outputs and few local plants that use either crude or refined palm oil as one of their inputs. This is not surprising, since mill location choice is driven by land suitability, rather than backward and forward linkages.

We study cohorts of palm oil mills between 2005 and 2015. Since we look at an event window including five time steps before the treatment year, and since we exclude plant-year observations that have seen previous treatment up to six years before, we benefit from the long manufacturing census panel and make use of manufacturing plants starting from 1994 to increase balance between our cohorts. We create a manufacturing census panel based on a manufacturing plant identifier that is consistently measured through the different survey waves and harmonize our outcome variables over those waves.

Many Indonesian districts have split in the study period, especially those on the outer islands outside of Java with natural resources and a history of ethnic conflict (Bazzi and Gudgeon 2018; Burgess et al. 2012; Pierskalla 2016). We therefore collapse districts into their polygon from 1993, which is the oldest year up to which BPS could provide geographical crosswalks.

During the study period, the manufacturing census has had a response rate between 65 (in 2011) and around 90% (in the 1990s).[21] BPS imputes missing values

[20]Sampling is done based on the Indonesian manufacturing directory, which includes the name, the number of workers, the addresses and contact information of all manufacturing plants. Budgets of field offices depend on the number of reporting establishments. They have an incentive to register new manufacturing plants (Blalock and Gertler 2004).

[21]For further background information refer to the annual print publication *Statistik Industri Manufaktur*, available from BPS or on request from the authors. This publication summarizes the findings from the manufacturing census.

based on previous periods and plants from the same industry: We detect and remove these duplicates. Indonesian law guarantees that manufacturing census data will not be used for other purposes than statistics. Still, some misreporting of financial information can be expected due to remaining concerns that data may be used by the government for tax collection or similar ends.

We obtained confidential data on the values and quantities of all inputs and all outputs of individual manufacturing plants between 1998 and 2015. Inputs and outputs are classified into 9-digit commodity codes in the Indonesian product classification system (KBKI/KKI), which is based on the international HS system. We use this data to construct indicator variables that capture whether a plant uses a new input or produces a new output in a given year. We also combine this plant-level data on outputs with classifications of tradability and relationship specificity. Our measure of tradability is based on the Holmes and Stevens (2014) classification.[22] We construct crosswalks from six-digit NAICS goods to Indonesian nine-digit commodity codes. We define goods as tradable if their η is lower than 0.8. We calculate the average plant-level share of tradable goods, weighted by the share of an individual input in the value of all inputs. We proceed in the same way for relationship specificity. The measure is based on the Rauch (1999) classification of goods.[23] We consider goods relationship-specific if they are neither goods traded on an organized exchange, nor goods with reference prices. We also use the detailed input and output data to clean our sales and total intermediate inputs variables.

For our estimation of revenue total factor productivity we rely on manufacturing plants' records on the book value of their machines, buildings, vehicles, and other capital. These fixed assets variables are missing in one third of our final sample. Further, we learned in our meetings with statistics officers in Jakarta that many plants do not properly record depreciated fixed assets. Therefore, we consistently compare total factor productivity to labor productivity in our main set of results. For our baseline specifications, we use the total of electricity bought from the grid and produced by a generator (in kWh) in the control function of our Levinsohn-Petrin production function estimator (Levinsohn and Petrin 2003) with Ackerberg-Caves-Fraser correction (Ackerberg et al. 2015). For robustness checks, we also use the total value of intermediate inputs (in Rp) in the control function. Since we have the list of all inputs and outputs on the plant-level we can check

[22]The Holmes classification of the tradability of goods can be accessed at `http://users.econ.umn.edu/~holmes/data/plantsize/index.html`

[23]The Rauch classification of the relationship-specificity of goods can be accessed at: `https://econweb.ucsd.edu/~jrauch/rauch_classification.html`

monetary values on individual items and compare their aggregate to the main intermediate inputs variable provided by the statistics office for plausibility. Lastly, we check robustness of our results (see Figure F.2) to different methods from the production function estimation literature (Ackerberg et al. 2015; Levinsohn and Petrin 2003; Wooldridge 2009).[24]

Figure 2 illustrates where our treatment variation comes from. We show tracts (in Indonesian *desas* or villages) that have manufacturing plants other than palm processing. Note that we use the district (borders printed in black) as our unit of observation in order to capture general equilibrium effects at the level of the local economy. Most palm oil districts have some tracts with manufacturing plants. Sumatra has more such tracts and higher numbers of plants, which reflects its longer history of industrialization.

Table 2 shows changes in our outcome variables between 2005 and 2015 for the full manufacturing census sample excluding the main island, Java, which has only two palm oil mills.

[Insert Table 2 here.]

District-level outcomes We also explore the impacts of palm oil mill construction on district-level outcomes. We are interested in district budgets (total taxes, natural resource revenues, forest revenues, national funds), infrastructure provision (roads spending, share of asphalt roads), and the local labor market (population size, total employment, employment in agriculture and industry, unemployment and poverty). We source these variables from a harmonized data set called INDO-DAPOER (Bank 2018). It is based on Indonesia's main surveys for employment (SAKERNAS), households (SUSENAS), and tracts (PODES). The data-set includes a crosswalk of districts, which we expand based on data provided by the Indonesian statistics agency, BPS, and which we expand and apply to the manufacturing census.

4 Results: Local agglomeration effects of palm oil mill shocks

Effects on manufacturing plant sales and productivity The establishment of a new palm oil mill increases sales and productivity of a district's manufacturing plants in comparison to plants in other districts that did not experience a palm oil

[24] See Appendix C.3.1 for more detail on the estimation of our production functions.

mill establishment over the preceding three years. Figure 4 illustrates dynamic effects on sales, labor productivity and total factor productivity. Dynamic effects level off after three to five years from treatment. In the five years prior to mill establishment, manufacturing plants from treatment and control districts show similar trends in all three outcomes.

[Insert Figure 4 here.]

Table 3 shows coefficients and standard errors from regressions of the natural log of our main set of outcome variables, sales, labor productivity and total factor productivity, on leads and lags of the establishment of a palm oil mill. The baseline specification defined in Equation 1 includes cohort-event time, cohort-treated, plant fixed effects, island-year fixed effects and industry-year fixed effects. Regressions using this specification are shown in columns 1, 4, and 7. If we include additional fixed effects (see remaining columns in the same table), our results show only small differences in magnitude and standard errors remain comparable. Pre-trends remain flat, when additional fixed effects are included.

[Insert Table 3 here.]

Table 4 shows estimates from Equation 2. These are difference-in-differences estimates for which we pool observations over the 5-year window after the establishment of the palm oil mill, excluding observations from year y, during which the mill is built. The establishment of a palm oil mill increases exposed manufacturing plants' sales by 15%, labor productivity by 13%, and total factor productivity by 13% compared to plants in palm districts without a shock three years before the event. We also investigate the extensive margin and find weak evidence for a small increase in plant creation and no evidence for increased plant closures (see Table E.4).

[Insert Table 4 here.]

We use the data on plant-level inputs and outputs to investigate if these spillover effects are due to linkages within supply chains to either upstream suppliers to, or downstream buyers from, palm oil processing facilities. Our sample contains only 125 treated plants downstream of palm oil processing facilities. This reflects the fact that palm oil resembles some extractive industries in which most of the downstream value is added in more distant locations. The most important downstream buyers of refined palm oil are producers of processed foods and cosmetics. Our sample contains 575 treated plants upstream. These produce fertilizer, building supplies, metalware, and chemicals used in the mill.[25] Dropping upstream and downstream plants from our baseline estimation sample does not substantially change our estimate; if anything, it leads to slightly larger effect size estimates (see Figure 5). Thus, we rule out that our effect is driven by positive spillovers on supplier or buyer plants, which in rare cases might even be part of the same firm. [26]. Furthermore, our results are robust to the exclusion of manufacturing plants that produce timber and pulp and paper products (see5). This indicates that our primary results are not driven by timber windfalls generated when forests are cleared for oil palm planting.

Labor market effects New palm oil mills and plantations create a shock to local labor markets. In the short-run, increases in the demand for plantation labor could increase blue collar wages. However, palm oil mills often actively encourage and support in-migration of laborers (Kelley et al. 2020). Furthermore, plantation establishment may restrict local communities' access to land for their own agricultural production (Li 2018). Increases in labor supply resulting from in-migration and a transition from subsistence to cash-crop agriculture could buffer wage increases. Given these counteracting dynamics, the aggregate wage effects of new palm oil mills are theoretically ambiguous.

In Table 5 we document effects on labor market outcomes, both at the plant and at the district level. Columns (1) to (3) are based on the manufacturing census sample and columns (4) to (12) use outcomes at the district-level from INDO-DAPOER. We estimate a 4% increase in blue-collar wages at the manufacturing plant-level, which is smaller than the increases in sales and productivity. We do not find any measurable increase in white-collar wages. This is not surprising, since palm oil mills and plantations create only a few office and engineering jobs. We do not find significant reductions in the number of workers per plant.

[Insert Table 5 here.]

We show increases in population size and employment at the district-level by approximately 50,000 and 20,000 people, respectively. The palm oil sector's reliance upon migrant labor could explain a portion of this uptick in population and employment (Li 2018). Underemployment also increases, which is consistent with

[25]Some mills also include food and fabrics for clothing in their supplies. These are likely provisions for workers, who are often also housed on the mill campus.

[26]Table E.5 shows estimates for the restricted sample of linked plants, which lack statistical power

the fact that palm oil plantations typically only provide season-dependent, part-time employment. Increased underemployment is also consistent with a transition from subsistence agriculture and labor intensive rubber and cocoa production towards less labor intensive oil palm cultivation. According to official statistics, smallholder agriculture makes up 40% of the planted oil palm area in Indonesia. Therefore, a large share of the estimated 2 million workers in Indonesia's palm oil sector are not formally employed by plantation companies, but engage with mills through markets for fresh fruit bunches (Qaim et al. 2020).[27] Based on the literature, we expect cash crop farming to increase agricultural productivity compared to subsistence farming (Qaim et al. 2020). However, it is unclear whether it frees up labor from farm work and thereby fosters industrial growth or whether it crowds out labor-market participation and non-farm entrepreneurship in contract farming households. Our findings speak to these research gaps with regards to spillovers on non-farm labor markets (Bellemare and Bloem 2018; Otsuka et al. 2016). We show that incumbent manufacturing plants are robust to potential crowding-out from increased wages on blue-collar labor markets. Our results complement previous findings based on household data, indicating that farmers' labor productivity increases by switching to palm oil, but they do not allocate more of their labor to employment (Euler et al. 2017).

Although oil palm cultivation is much more labor intensive than some forms of natural resource extraction, including oil, gas, and mining, our results mirror earlier findings that resource booms can increase manufacturing wages, while simultaneously encouraging manufacturing plants to increase labor productivity (Cust et al. 2019). We document wage increases that are double what Cust et al. (2019) find for a 10 % increase in oil and gas windfalls. Large increases in population size and employment in reaction to palm oil mill establishment point to substantial in-migration, which offsets part of an expected wage increase in the absence of labor mobility. We address some potential concerns around SUTVA violations due to migration from control districts into treated districts by merging city districts with rural districts for robustness checks (see Section 4.1).

With our stacked design, we also replicate the earlier result from long-difference and instrumental variable regressions, that palm oil plantations have decreased poverty in Indonesia (Edwards 2019a).

[27]Historically this has been driven by government interventions that required palm companies to share their concessions with local communities, but, today, there is an increasing share of independent smallholders planting oil palms on the edges of corporate concessions.

Road infrastructure and district budgets Government revenues and the quality of local transportation infrastructure improved after the establishment of new palm oil mills. Table 6 documents the impacts of palm oil mill establishment on district-level GDP, local government revenues, and on road investments and quality. We find that the establishment of a palm oil mill is associated with increases in local tax and resource revenues, including timber levies that are due when plantations are developed.

[Insert Table 6 here.]

Although government revenues increased substantially, we cannot measure strong increases in district infrastructure spending and national government expenditures on the roads within a district after the establishment of a mill. Nevertheless, new palm oil mills were associated with increases in the share of a district's tracts (*desa*) with asphalt roads. These seemingly contradictory results are consistent with the fact that, during the *laissez-faire* period of plantation development studied here, plantation firms were responsible for establishing necessary infrastructure without direct state support. The private infrastructure investments of oil palm firms appears to have driven significant upgrades in the road network at the tract-level (Gatto et al. 2017; McCarthy 2010).

National funds for agriculture increase in areas with new palm oil mill investments, hinting at a lower net effect of spillovers in the absence of government intervention. Robustness checks with region-time fixed effects at the geographical level just above our treatment variation (see Figure 5) do not change the magnitude of our main effects (reported in Table 4). This means that higher-level investments by the national government, such as provincial roads, highways or ports cannot be the main driver of local spillover effects.

Increases in district agricultural and industrial GDP (Table 4, Columns 1-2) are consistent with previous findings on the impacts of palm oil mills on surrounding village economies. Using the cross-section of palm oil mills, Edwards (2019b) documents higher employment, more plants and improved public goods, such as roads, markets and public transport, in villages within a 20 kilometer radius. Our results indicate that economic spillovers from palm oil plantations on manufacturing are significant at the larger geographical level of the district economy.

Product portfolio effects After the establishment of a new palm oil mill within a district, nearby manufacturing plants shift their product portfolio towards tradable goods, while decreasing the share of relationship-specific goods. Table 7 presents evidence for these changes in manufacturing product portfolios. Observed increases in the share of tradable goods (2%) are consistent with improvements in transport infrastructure due to palm oil booms. Decreases in the share of relationship-specific-goods (-1%) could reflect the same shift from supplying inputs to other plants in the same region towards producing for supra-regional markets caused by better transport infrastructure.

These shifts in product portfolio warrant further investigation, since they document changes in production functions that could have important implications for productivity in the mid-term by changing learning-by-doing dynamics or investments in innovation.

[Insert Table 7 here.]

4.1 Robustness checks

Our stacked research design allows us to exercise more control over treatment and control units for each cohort. This also creates additional researcher degrees of freedom, which cannot be exhaustively reflected by standard regression tables.

We therefore build specification charts (see Figures 5 and Figures F.1, and F.2 in the Online Appendix) that compare point estimates from standard dose-response fixed effects regressions with stacked specifications under different corporate group samples, event-window sizes, control group definitions, assumptions on anticipation, and data sources. These charts also include coefficients for different combinations of fixed effects. We include fixed effects on the sector-island-year level to absorb idiosyncratic shocks to industries in specific islands that may be driving treatment adoption and outcomes. We include specifications with eight-digit industry-time and province-year fixed effects. Provinces are the geographical unit just above districts, where our treatment varies.

[Insert Figure 5 here.]

Estimates of a standard two-way fixed effects panel regression with the count of mills in each district as the treatment variable (see Figure 5) show effect sizes of similar magnitudes. Goodman-Bacon (2018) shows that any difference could come from two sources: either common trends violations or the weighted aggregation of heterogeneous effects between cohorts and between groups of different treatment.[28]

[28]See Appendix B.1 for a detailed discussion of the difference between our stacked design and standard dose-response two-way fixed effects specifications.

We elaborate on robustness checks, examining our research design's main remaining threats to identification in the following paragraphs.[29]

Restricting sample to large corporate groups In our baseline specification, we use the full sample of Indonesian palm oil mills to define treatment cohorts. A key identifying assumption of our research is that, conditional on fixed effects, treatment is assigned independently from local shocks. However, small companies operating in a restricted geographic area may make mill establishment decisions in response to local shocks. To test the robustness of our results to this concern, we re-run our primary analysis using restricted samples consisting only of mills belonging to corporate groups that control large land banks with multiple licensed but undeveloped concessions. These corporate groups base their new investment decisions on factors that are either time-invariant or vary at the level of our time-varying fixed effects (see Section 2.2 for a detailed argumentation). We look at three different samples: (i) restricted to groups with at least 5 mills, (ii) restricted to groups above 25 mills ("big five"), and (iii) restricted to groups above 25 mills, but without the state-owned company PTPN III. As shown in Table E.1, palm oil groups have diversified their interests over the main Indonesian palm oil islands, Sumatra and Kalimantan. We show that our main result stays robust when estimated in these samples, but is attenuated as the sample becomes more restricted.

Restricting sample to palm oil areas In our baseline specifications we only compare manufacturing plants on the same island and in the same industry. In addition, we run robustness checks with province-year fixed effects. However, even when comparing manufacturing plants within a single island or province, areas without oil palm may be exposed to different time-varying factors than oil palm-growing areas, since they are more urban or topographically different. Therefore, in robustness checks we also exclude districts that were never treated. Coefficients stay positive and significant for this smaller sample, but are attenuated. The results show that, in existing palm oil areas, incumbent non-palm oil manufacturing plants benefit from the establishment of an additional palm oil mill.

Spillovers to neighboring districts The construction of a mill in one district can impact neighboring districts with pre-existing mills or mills planned in the

[29] See Appendix A.

future, and thereby bias our estimates. Our main specification only uses year-observations in the control group that are not from a manufacturing plant exposed to treatment in the same year and the three preceding years. Still, the stable unit treatment value assumption (SUTVA) may be violated for these controls, since a new mill might have local effects on input and output prices, labor markets and intra-company resource allocation. This is of particular concern for the sample with more regional rather than national and global companies. Bias could run in both directions: new mills could affect incumbent mills negatively by lowering palm oil prices or by diverting workers and investment, but they could also affect them positively by improving their parent company's financial situation.

We estimate spillovers to neighboring districts, often palm oil districts themselves, excluding the treatment districts from the regression (see Table 8). The coefficient estimates are small in magnitude and statistically insignificant (see Table 2). The lack of evidence for spillovers in this setup supports the tenability of the SUTVA.

[Insert Table 8 here.]

Never-treated controls only We also check robustness to a broader type of SUTVA violation by restricting controls to never-treated districts. We find higher treatment effects when comparing treated districts to these non-palm oil districts. Note that never-treated districts on the same island are typically either more urbanized or are unsuitable for plantations. Therefore this difference in effects could, in part, reflect endogeneity. Nevertheless, this robustness test suggests that spillovers likely attenuate our estimated treatment effects, rather than bias them upwards.

Including cities Since Indonesian cities are categorized as their own districts, our baseline specification does not capture spillovers to manufacturing plants that are located in city clusters, but benefit from the shock to their surrounding more rural areas. We therefore merge those cities that are surrounded by rural districts with those rural districts and run specifications on the sample of districts that have palm oil, effectively still excluding those cities that cannot be assigned to a rural district. Coefficients from regressions on this sample are smaller than our baseline regressions. This attenuation could be a result of the fact that city economies experience different economic shocks than palm oil-based rural economies.

Changing event and exclusion windows Our baseline specification compares a window of five years before treatment to a window of five years after treatment and excludes observations from the control groups three years after a manufacturing plant gets treated. These a priori choices are based on the functioning of the palm oil economy. However, we also check robustness to changing these parameters of control group choice. We run robustness checks with longer event windows of eight years, which capture some dynamics beyond our standard five years window. We also adapt the exclusion window to six years post-treatment exclusion and three years pre-treatment exclusion. We show that changing these parameters only leads to small changes in magnitudes and standard errors, with some specifications leading to slightly higher and some to slightly smaller coefficients.

Excluding anticipation years from event window Anticipation effects for palm oil mills can arise since palm plants have to wait for their oil palms to grow. There is a typical gap of three years between the planting of oil palms and the first harvest, when the mill has to be established. During this phase, workers are needed to plant palms and to start building roads. These activities are highly visible and will create expectations in incumbent plants and workers. Therefore, the district economy may begin its adjustment before the date palm oil production starts. Excluding three periods before treatment in robustness checks leads to coefficients with similar magnitudes, but increases standard errors.

5 Discussion

In this paper, we provide evidence for spillovers from investments in palm oil to incumbent non-palm oil manufacturing plants that do not run directly through supply chains. The manufacturing census provides us with detailed information on all manufacturing plants in Indonesia, also in rural areas outside of the main industrial centers of the country. Our panel of mills allows us to construct plausible control groups for individual event cohorts and unpack the dynamic effects of palm oil shocks on non-palm oil manufacturing plants. Our point estimates of the average agglomeration effect after the construction of a new palm oil mill are 15% for sales, 13% for labor productivity, and 13% for total factor productivity. Blue collar wages increase by 4%, indicating that there is some competition for labor between industry and agriculture, but this increase is attenuated by in-migration.

We also see changes in manufacturing plant production patterns. They reduce their share of relationship-specific goods, but increase their share of tradable goods. This change in the manufacturing sector's production portfolio may reflect a response to the 12% percent increase in the share of a district's tracts with paved roads resulting from the establishment of a new palm oil mill. Finally, new mills reduce poverty and increase tax returns for local governments.

Study limitations The average treatment effect we estimate leverages comparisons between plants that are all based in palm oil agglomeration areas or at least on the same island. This raises concerns about the external validity of our results.

First, we cannot answer what would have happened to these areas if the oil palm had never been imported there from West Africa. Since data on the Indonesian palm oil sector is hard to obtain, we mainly have clean mill establishment dates for the most recent phase of the plantation expansion. This means we cannot construct a counterfactual based on a comparison between treated and not yet treated districts in the early phase of the palm oil boom in the 1980s and 1990s. We cannot rule out that a crowding-out has happened in that period and that our study only speaks to the effects in a restricted sample, which is more resilient to crowding out due to a prior reshuffling with attrition from and selection into palm areas.

Second, since the Indonesian government put political and budgetary resources into palm areas in parallel to corporate investments in palm oil, we cannot answer what the net effects of palm oil mills without government intervention would have been. For instance, governments invest in repairing roads. The Public Works Office of a provincial government in Kalimantan on the island of Borneo estimated in 2006 that more than half of its roads were in poor condition, due to trucks with heavy loads (Public Works of Central Kalimantan Province 2006). Governments might also have neglected other islands in order to fund the palm oil expansion. We also cannot say, whether the Indonesian government could have created more industrial growth if it had supported alternative economic activities in the concerned regions.

Third, we only capture dynamic effects up to five years after treatment. While lags of our coefficient indicate that agglomeration spillovers level off within this study window, there might be dynamics that only surface in the mid or long run. For instance, Coxhead and Shrestha (2016) find that intensity in palm oil production is linked to lower formal employment, which is a key driver of investments in education.

Further research Our estimated effects also do not capture effects that the Indonesian palm oil boom had on the national level. Our study only uses variation within manufacturing plants in the same region and in the same industry. Therefore, effects that are common to the palm oil sector as a whole, to all sectors, or to all regions do not appear in our effects. Many channels that have been discussed in the natural resource curse literature act in general equilibrium and at the national level. The empirical challenge of finding a suitable counterfactual for the Indonesian palm oil boom is similar to the case of colonial sugar factories examined by Dell and Olken (2020). Villages with colonial sugar factories and plantations are more developed today than similar nearby villages without them. However, this counterfactual does not tell us whether Indonesia as a whole is more developed today than it would have been without sugar plantations. Similarly, our study cannot answer whether the Indonesian manufacturing share in GDP would be higher today without the expansion of palm oil plantations in its outer islands. Whether there is a palm oil resource curse should, therefore, be answered by studies that use the island or the country as their unit of observation.

Policy relevance The Indonesian palm oil boom has coincided with a phase of slower industrialization. As governments consider policies to support both agricultural and industrial development, it is important to know whether the rapid growth of the agricultural sector has crowded out industrial activity in this region. To answer this empirical question, a credible counterfactual has to be constructed. The placement of palm oil mills is endogenous to growth prospects, infrastructure planning and other time-varying unobservables. Also, relatively few districts in palm-suitable regions have not seen the establishment of palm oil mills and these are typically more urban and better connected areas.

We leverage variation in the timing and location of palm oil mill establishment to assess production changes in manufacturing plants in other sectors. We provide evidence that incumbent plants resist crowding-out and even benefit from new palm oil booms, as we detect positive local spillovers on industrial development. However, this does not mean that a palm oil-based development strategy has been optimal for Indonesia. This question can only be answered by a full cost-benefit analysis that takes into account a comprehensive set of social and environmental costs and benefits, including a counterfactual of the national economy in general equilibrium.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83.6, pp. 2411–2451.

Allcott, Hunt and Daniel Keniston. 2018. "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America". *The Review of Economic Studies* 85.2, pp. 695–731.

Athey, Susan and Guido Imbens. 2018. "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption". arXiv: `1808.05293 [cs, econ, math, stat]`.

Badan Pusat Statistik (BPS). 2018. *Survei Industri Besar/Sedang (IBS): 1994-2015*. URL: `https://mikrodata.bps.go.id/mikrodata/index.php/catalog/IBS` (visited on 04/24/2018).

Bank, The World. 2018. *Indonesia Database for Policy and Economic Research (INDO-DAPOER)*. URL: `https://datacatalog.worldbank.org/dataset/indonesia-database-policy-and-economic-research` (visited on 08/02/2018).

Baudoin, Alice et al. 2019. *Review of the Diversity of Palm Oil Production Systems in Indonesia - Case Study of Two Provinces: Riau and Jambi*. Working Paper 219. Bogor, Indonesia: Center for International Forestry Research (CIFOR).

Bazzi, Samuel and Matthew Gudgeon. 2018. *The Political Boundaries of Ethnic Divisions*. Working Paper 24625. Series: Working Paper Series. National Bureau of Economic Research.

Bellemare, Marc F. and Jeffrey R. Bloem. 2018. "Does Contract Farming Improve Welfare? A Review". *World Development* 112, pp. 259–271.

Blalock, Garrick and Paul J. Gertler. 2004. "Learning from exporting revisited in a less developed setting". *Journal of Development Economics*. 15th Inter American Seminar on Economics 75.2, pp. 397–416.

Burgess, Robin et al. 2012. "The Political Economy of Deforestation in the Tropics". *The Quarterly Journal of Economics* 127.4, pp. 1707–1754.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli. 2016. "Agricultural Productivity and Structural Transformation: Evidence from Brazil". *American Economic Review* 106.6, pp. 1320–1365.

Byerlee, Derek, Walter P. Falcon, and Rosamond Naylor. 2016. *The Tropical Oil Crop Revolution: Food, Feed, Fuel, and Forests*. Oxford University Press.

Cengiz, Doruk et al. 2019. "The Effect of Minimum Wages on Low-Wage Jobs". *The Quarterly Journal of Economics* 134.3, pp. 1405–1454.

Corley, R.H.V. and P.B. Tinker. 2016. *The Oil Palm*. Fifth Edition. Chichester, UK: John Wiley & Sons, Ltd.

Coxhead, Ian and Rashesh Shrestha. 2016. "Could a Resource Export Boom Reduce Workers' Earnings? The Labour-Market Channel in Indonesia". *Bulletin of Indonesian Economic Studies* 52.2, pp. 185–208.

Cust, James, Torfinn Harding, and Pierre-Louis Vézina. 2019. "Dutch Disease Resistance: Evidence from Indonesian Firms". *Journal of the Association of Environmental and Resource Economists* 6.6, pp. 1205–1237.

Cust, James and Steven Poelhekke. 2015. "The local economic impacts of natural resource extraction". *Annu. Rev. Resour. Econ.* 7.1, pp. 251–268.

De Chaisemartin, Clément and Xavier D'HaultfŒuille. 2018. "Fuzzy Differences-in-Differences". *The Review of Economic Studies* 85.2, pp. 999–1028.

Dell, Melissa and Benjamin A. Olken. 2020. "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java". *The Review of Economic Studies* 87.1, pp. 164–203.

Deshpande, Manasi and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs". *American Economic Journal: Economic Policy* 11.4, pp. 213–248.

Edwards, Ryan B. 2019a. "Export Agriculture and Rural Poverty: Evidence from Indonesian Palm Oil". Unpublished Manuscript.

— . 2019b. "Spillovers from Agricultural Processing". Unpublished Manuscript.

Euler, Michael et al. 2017. "Oil Palm Adoption, Household Welfare, and Nutrition Among Smallholder Farmers in Indonesia". *World Development* 93, pp. 219–235.

Faber, Benjamin and Cecile Gaubert. 2019. "Tourism and Economic Development: Evidence from Mexico's Coastline". *American Economic Review* 109.6, pp. 2245–2293.

Foster, Andrew D. and Mark R. Rosenzweig. 2004. "Agricultural productivity growth, rural economic diversity, and economic reforms: India, 1970–2000". *Economic Development and Cultural Change* 52.3, pp. 509–542.

— . 2007. "Economic development and the decline of agricultural employment". *Handbook of development economics* 4, pp. 3051–3083.

Gatto, Marcel, Meike Wollni, and Matin Qaim. 2015. "Oil Palm Boom and Land-Use Dynamics in Indonesia: The Role of Policies and Socioeconomic Factors". *Land Use Policy* 46, pp. 292–303.

Gatto, Marcel et al. 2017. "Oil Palm Boom, Contract Farming, and Rural Economic Development: Village-Level Evidence from Indonesia". *World Development* 95, pp. 127–140.

Goodman-Bacon, Andrew. 2018. *Difference-in-Differences with Variation in Treatment Timing*. Working Paper 25018. National Bureau of Economic Research.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings". *Journal of Political Economy* 118.3, pp. 536–598.

Heilmayr, Robert, Kimberly M. Carlson, and Jason Jon Benedict. 2020. "Deforestation Spillovers from Oil Palm Sustainability Certification". en. *Environmental Research Letters*.

Holmes, Thomas J. and John J. Stevens. 2014. "An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade". *Journal of Political Economy* 122.2, pp. 369–421. `http://users.econ.umn.edu/~holmes/data/plantsize/index.html`.

Jelsma, Idsert et al. 2017. "Unpacking Indonesia's Independent Oil Palm Smallholders: An Actor-Disaggregated Approach to Identifying Environmental and Social Performance Challenges". *Land Use Policy* 69, pp. 281–297.

Kelley, Lisa C. et al. 2020. "Circular Labor Migration and Land-Livelihood Dynamics in Southeast Asia's Concession Landscapes". en. *Journal of Rural Studies* 73, pp. 21–33.

Kubitza, Christoph and Esther Gehrke. 2018. *Why Does a Labor-Saving Technology Decrease Fertility Rates? Evidence from the Oil Palm Boom in Indonesia*. EFForTS Discussion Paper Series.

Levinsohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables". *The Review of Economic Studies* 70.2, pp. 317–341.

Li, Tania Murray. 2018. "After the Land Grab: Infrastructural Violence and the "Mafia System" in Indonesia's Oil Palm Plantation Zones". *Geoforum* 96, pp. 328–337.

McCarthy, John F. 2010. "Processes of Inclusion and Adverse Incorporation: Oil Palm and Agrarian Change in Sumatra, Indonesia". *The Journal of Peasant Studies* 37.4. _eprint: https://doi.org/10.1080/03066150.2010.512460, pp. 821–850.

Moscona, Jacob. 2019. "Agricultural Development and Structural Change Within and Across Countries". Unpublished Manuscript.

Otsuka, Keijiro, Yuko Nakano, and Kazushi Takahashi. 2016. "Contract Farming in Developed and Developing Countries". *Annual Review of Resource Economics* 8.1, pp. 353–376.

Pierskalla, Jan H. 2016. "Splitting the Difference? The Politics of District Creation in Indonesia". *Comparative Politics* 48.2, pp. 249–268.

Public Works of Central Kalimantan Province, Office of. 2006. *Sinergi Antara Transportasi Jalan Dan Sungai Di Kalimantan Tengah Menuju Sistem Berkelanjutan, Palangka Raya*.

Qaim, Matin et al. 2020. "Environmental, Economic, and Social Consequences of the Oil Palm Boom". en. *Annual Review of Resource Economics* 12.1.

Rauch, James E. 1999. "Networks versus Markets in International Trade". *Journal of International Economics* 48.1, pp. 7–35. https://econweb.ucsd.edu/~jrauch/rauch_classification.html.

The Jakarta Post. 2016. *Commodity Boom Causes Premature Deindustrialization*. The Jakarta Post. URL: https://www.thejakartapost.com/news/2016/09/15/commodity-boom-causes-premature-deindustrialization.html (visited on 12/01/2020).

Wooldridge, Jeffrey M. 2009. "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables". *Economics Letters* 104.3, pp. 112–114.

World Resources Institute et al. 2019. *Universal Mill List*. URL: https://data.globalforestwatch.org/datasets/5c026d553ff049a585b90c3b1d53d4f5_34 (visited on 12/14/2020).

Tables

Table 1: Number of treated and control districts for each cohort in the stacked dataset

Cohort	Treated	Control
2005	17	258
2006	15	268
2007	9	277
2008	12	273
2009	15	268
2010	14	264
2011	17	261
2012	18	261
2013	18	261
2014	17	261
2015	16	259

Notes. This table reports the number of districts in the treatment and control group in the stacked sample for each cohort of palm oil mills from 2005 to 2015. It is based on data on the establishment of mills collected from provincial plantation offices.

Table 2: Summary statistics manufacturing plants

	2005 Mean	Median	Std. Dev.	2015 Mean	Median	Std. Dev.
Firm performance:						
Total Factor Productivity (LP ACF, electricity, log)	7.78	7.74	1.14	8.19	8.19	1.26
Sales (in 10,000 USD)	1166.8	56.5	6618.8	1339.1	91.6	11485.6
Labor:						
Number of workers	203.3	46	550.1	169.8	47	526.4
Annual wage blue-collar workers (in USD)	1533.5	1329.0	1173.0	1921.0	1796.4	1029.4
Annual wage white-collar workers (in USD)	2477.4	1591.2	2965.4	2431.0	1997.2	1919.8
Labor productivity (output per worker in USD)	31474.7	11075.2	80785.2	77190.9	16940.8	1454775.5
Inputs:						
Number of inputs used by plant	3.72	3	2.83	2.84	1	3.98
Imported materials (in 10,000 USD)	504.8	0	4815.6	926.7	0	8515.3
Domestic materials (in 10,000 USD)	2804.0	97.0	20099.1	5206.0	292.1	40450.8
Electricity consumption (MWh)	3357.8	106.7	28308.9	2121.3	61.7	21754.1
Product portfolio:						
Number of outputs produced by plant	1.86	1	1.50	1.59	1	1.26
Share of tradable goods in outputs	0.46	0.41	0.39	0.51	0.50	0.39
Share of relationship-specific goods in outputs	0.53	0.74	0.48	0.51	0.67	0.49

Notes. This table reports summary statistics for outcomes and auxiliary variables (for production function estimations) at the manufacturing plant level. Data is from the Indonesian manufacturing census, which includes a total of 18071 plants in 1994, 19887 in 2005, and 22416 in 2015 (excluding palm oil plants). We report mean, median, and standard variation for these variables in 2005 and 2015, which are the starting and end years of our study period, excluding observations from the island Java, from where no identification variation comes. Monetary amounts are in 2010 USD. The share of tradable goods in manufacturing goods is calculated based on the share of the value of an individual output in the value of all outputs after categorizing outputs according to the classification by Holmes and Stevens (2014). We calculate the share of relationship-specific goods in the same manner, based on Rauch (1999).

Table 3: Leads and lags of palm oil mill establishment on local manufacturing plant performance

	Sales (log)			Labor productivity (log)			Total Factor Productivity (log)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mill est. (t+5)	−0.007	−0.011	−0.014	0.001	−0.003	−0.007	−0.027	−0.041	−0.051
	(0.038)	(0.037)	(0.032)	(0.032)	(0.032)	(0.028)	(0.037)	(0.035)	(0.028)
Mill est. (t+4)	−0.014	−0.017	−0.012	−0.008	−0.013	−0.009	−0.011	−0.022	−0.019
	(0.026)	(0.025)	(0.020)	(0.022)	(0.022)	(0.020)	(0.032)	(0.031)	(0.026)
Mill est. (t+3)	−0.014	−0.016	−0.014	−0.009	−0.013	−0.010	−0.019	−0.028	−0.028
	(0.024)	(0.023)	(0.016)	(0.022)	(0.022)	(0.016)	(0.023)	(0.022)	(0.019)
Mill est. (t+2)	−0.022	−0.026	−0.031	−0.022	−0.027	−0.032	−0.009	−0.019	−0.026
	(0.016)	(0.015)	(0.013)	(0.017)	(0.016)	(0.013)	(0.025)	(0.024)	(0.021)
Mill est. (t)	0.023	0.026	0.013	0.017	0.018	0.005	0.097	0.094	0.053
	(0.025)	(0.024)	(0.017)	(0.029)	(0.028)	(0.022)	(0.028)	(0.027)	(0.027)
Mill est. (t-1)	0.116	0.114	0.107	0.098	0.096	0.091	0.133	0.131	0.084
	(0.031)	(0.031)	(0.032)	(0.029)	(0.028)	(0.027)	(0.034)	(0.033)	(0.031)
Mill est. (t-2)	0.132	0.128	0.106	0.115	0.107	0.074	0.104	0.096	0.057
	(0.030)	(0.029)	(0.035)	(0.030)	(0.029)	(0.031)	(0.033)	(0.032)	(0.031)
Mill est. (t-3)	0.140	0.138	0.154	0.127	0.122	0.130	0.103	0.097	0.082
	(0.039)	(0.038)	(0.043)	(0.036)	(0.036)	(0.035)	(0.042)	(0.045)	(0.037)
Mill est. (t-4)	0.145	0.139	0.150	0.128	0.121	0.117	0.157	0.148	0.111
	(0.045)	(0.045)	(0.050)	(0.041)	(0.040)	(0.038)	(0.037)	(0.037)	(0.035)
Mill est. (t-5)	0.157	0.151	0.142	0.137	0.127	0.099	0.174	0.165	0.124
	(0.047)	(0.045)	(0.051)	(0.043)	(0.040)	(0.039)	(0.054)	(0.051)	(0.048)
Cohort-event time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort-treated FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Island-year FE	Y	Y		Y	Y		Y	Y	
Industry-year FE	Y	Y		Y	Y		Y	Y	
Industry-island FE		Y			Y			Y	
Industry-island-year FE			Y			Y			Y
District clusters	285	285	285	285	285	285	283	283	283
N	2068029	2068018	2067810	2068029	2068018	2067810	1323191	1323179	1322985

Notes. This table reports the dynamic effects of a new palm oil mill on non-palm oil manufacturing plant performance. These are the coefficients from Equation 1. They are also shown in Figure 4. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses.
The unit of observation in this sample is the manufacturing plant. We have yearly observations. Plants are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time, rather than calendar time. We exclude observations from cohorts if they come from manufacturing plants that were treated within five year before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample).
Columns (1)–(3) show coefficients for the natural log of sales as the outcome, columns (4)–(6) for the natural log of labor productivity (sales per worker), column (7)–(9) for the natural log of revenue total factor productivity estimated with the Levinsohn-Petrin estimator with Ackerberg-Caves-Fraser correction and electricity consumption as the instrument. Columns (1), (4), and (7) include our baseline set of fixed effects, including cohort-event time FE, cohort-treated FE, firm FE, island-year FE, and industry-year FE (at the five-digit sector level defined by a plant's main output). Columns (2), (5), and (8) include an additional industry-island FE. Columns (3), (6), and (9) include an industry-island-year FE, which absorbs the island-year and industry-year FE.

Table 4: Effects of palm oil mill establishment on local manufacturing plant performance

	(1) Sales (log)	(2) Labor prod. (log)	(3) TFP (log)
Mill est. (t-5,t-1)	0.145	0.125	0.136
	(0.036)	(0.031)	(0.033)
Cohort-event time FE	Y	Y	Y
Cohort-treated FE	Y	Y	Y
Firm FE	Y	Y	Y
Island-year FE	Y	Y	Y
Industry-year FE	Y	Y	Y
District clusters	285	285	283
N	1851041	1851041	1187522

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on non-palm oil manufacturing plant performance. These are the coefficients from our baseline Equation 2.
The unit of observation in this sample is the manufacturing plant. Observations are pooled over the five years before and after a mill is established. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses. We have yearly observations.
In the estimation sample, manufacturing plants are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time rather than calendar time. We exclude observations from cohorts if they come from manufacturing plants that were treated within five years before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample).
Column (1) shows the coefficient for the natural log of sales as the outcome, column (2) for the natural log of labor productivity (sales per workers), column (3) for the natural log of revenue total factor productivity, estimated with the Levinsohn-Petrin estimator with Ackerberg-Caves-Fraser correction and electricity consumption as the instrument. All three columns include our baseline set of fixed effects, i.e., cohort-event time FE, cohort-treated FE, firm FE, island-year FE, and industry-year FE (at the five-digit sector level defined by a plant's main output).

Table 5: Effects on population, employment, and poverty at manufacturing plant and district levels

	Manufacturing plant sample			District sample						
	(1) Wage blue collar (log)	(2) Wage white collar (log)	(3) Workers in plants (log)	(4) Population (in 1000)	(5) Employed (in 1000)	(6) Employed agriculture (in 1000)	(7) Employed industry (in 1000)	(8) Under employed (in 1000)	(9) Poor persons (in 1000)	(10) Poverty rate (log)
Mill est. (t-5,t-1)	0.043	0.017	0.020	64.278	16.328	5.897	0.192	9.440	−15.379	−0.146
	(0.016)	(0.024)	(0.022)	(23.163)	(7.924)	(3.716)	(1.001)	(2.955)	(4.988)	(0.037)
Cohort-event time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort-treated FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y							
District FE				Y	Y	Y	Y	Y	Y	Y
Island-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y							
District clusters	285	285	285	280	279	279	279	279	280	280
N	1803569	1477295	1851041	47590	18949	18885	18870	18949	32728	32728

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on non-palm oil manufacturing plant labor outcomes and district-level indicators. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses.
The unit of observation for columns (1) to (3) is the manufacturing plant and data comes from the manufacturing census. For columns (4) to (10), it is the district, and data comes from the INDO-Dapoer data. For both samples, we have yearly observations. Plants and districts are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time rather than calendar time. We exclude observations from cohorts if they come from units that were treated within five years before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample).
Column (1) reports effects on the natural log of blue collar wages (often "unskilled labor" in the literature), column (2) on the natural log of white collar wages ("skilled labor"). Both variables are the annual wage bill divided by the number of workers in the respective categories. Column (3) shows the effect on the natural log of the number of workers per plant.
Column (4) reports effects on the total population of a district, (5) the number of employed people, (6) the number of people employed in agriculture (note that this constitutes the bulk of palm oil labor), (7) the number of people employed in industry, (8) the number of underemployed people, (9) the number of poor persons per district, and (10) the natural log of the poverty rate.
Coefficients from all specifications include our cohort-event time FE, cohort-treated FE, firm FE, and island-year FE. Regressions at the manufacturing plant level also include industry-year FE (at the five-digit sector level defined by a plant's main output).

Table 6: Effects on district budgets and infrastructure

	(1) GDP Agriculture (log)	(2) GDP Manufacturing (log)	(3) District tax revenue (log)	(4) Natural resource revenues (log)	(5) Forestry revenues (log)	(6) National funds roads (log)	(7) National funds agriculture (log)	(8) District expenditure infrastructure (log)	(9) Asphalt roads (share log)
Mill est. (t-5,t-1)	0.050	0.082	0.186	0.398	0.399	0.045	0.144	0.078	0.112
	(0.021)	(0.036)	(0.046)	(0.139)	(0.284)	(0.076)	(0.081)	(0.082)	(0.030)
Cohort-event time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort-treated FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Island-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
District clusters	281	281	275	275	266	275	274	275	280
N	34142	34142	43324	42054	12021	16020	10761	27741	15921

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on district-level outcomes. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses.
The unit of observation is the district and data comes from the INDO-Dapoer data. We have yearly observations. Plants and districts are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time rather than calendar time. We exclude observations from cohorts if they come from units that were treated within five years before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample).
Columns (1) and (2) report effects on GDP in agriculture and manufacturing respectively. Column (3) reports effects on overall tax revenues at the district level. Column (4) reports effects on natural resources revenues and column (5) reports effects on forestry revenues. Column (6) reports funds received by a district from the national government for the construction of roads and column (7) reports the same type of funding but for agricultural activities. Column (8) reports expenditure in infrastructure by the district government and column (9) reports the share of asphalt roads in the tracts (in Bahasa Indonesia *desas*) of a district. All outcome variables are included with their natural logs.
Coefficients from all specifications include our cohort-event time FE, cohort-treated FE, firm FE, and island-year FE.

Table 7: Effects on manufacturing plant output portfolio

	(1) Tradable share (log)	(2) Specific share (log)
Mill est. (t-5,t-1)	0.021	−0.010
	(0.011)	(0.007)
Cohort-event time FE	Y	Y
Cohort-treated FE	Y	Y
Firm FE	Y	Y
Island-year FE	Y	Y
Industry-year FE	Y	Y
District clusters	273	285
N	573202	1839614

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on non-palm oil manufacturing plant performance. These are the coefficients from our baseline Equation 2. The unit of observation in this sample is the manufacturing plant. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses. We have yearly observations. Column (1) reports on the natural log of the share of tradable products in a plant's outputs according to the categorization by Holmes and Stevens (2014) and (2) reports on the the natural log of the share of relationship-specific products in a plant's outputs according to the categorization by Rauch (1999). Both columns include our baseline set of fixed effects, i.e., cohort-event time FE, cohort-treated FE, firm FE, island-year FE, and industry-year FE (at the five-digit sector level defined by a plant's main output).

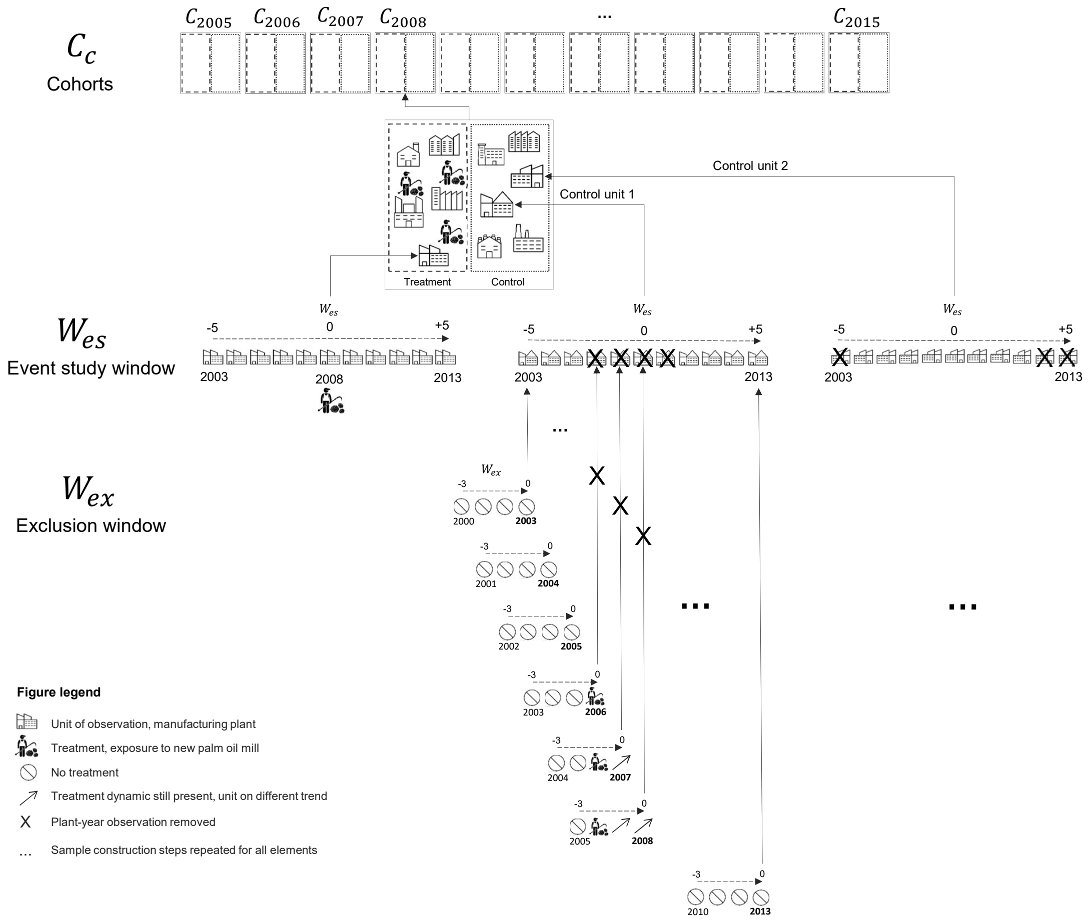
Table 8: Checking SUTVA - Effects of palm oil mill establishment on neighboring districts

	(1) Sales (log)	(2) Labor prod. (log)	(3) TFP (log)
Mill est. (t-5,t-1)	−0.021	−0.019	0.053
	(0.034)	(0.035)	(0.046)
Cohort-event time FE	Y	Y	Y
Cohort-treated FE	Y	Y	Y
Firm FE	Y	Y	Y
Island-year FE	Y	Y	Y
Industry-year FE	Y	Y	Y
District clusters	284	284	282
N	1809111	1809111	1160143

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on non-palm oil manufacturing plant performance in neighboring districts. The specification is from our baseline Equation 2. The unit of observation in this sample is the manufacturing plant. Observations are pooled over the five years before and after a mill is established. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses. We have yearly observations. In the estimation sample, manufacturing plants are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time rather than calendar time. We exclude observations from cohorts if they come from manufacturing plants that were treated within five years before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample). Column (1) shows the coefficient for the natural log of sales as the outcome, column (2) for the natural log of labor productivity (sales per worker), column (3) for the natural log of revenue total factor productivity, estimated with the Levinsohn-Petrin estimator, with Ackerberg-Caves-Fraser correction and electricity consumption as the instrument. All three columns include our baseline set of fixed effects, i.e., cohort-event time FE, cohort-treated FE, firm FE, island-year FE, and industry-year FE (at the five-digit sector level defined by a plant's main output).

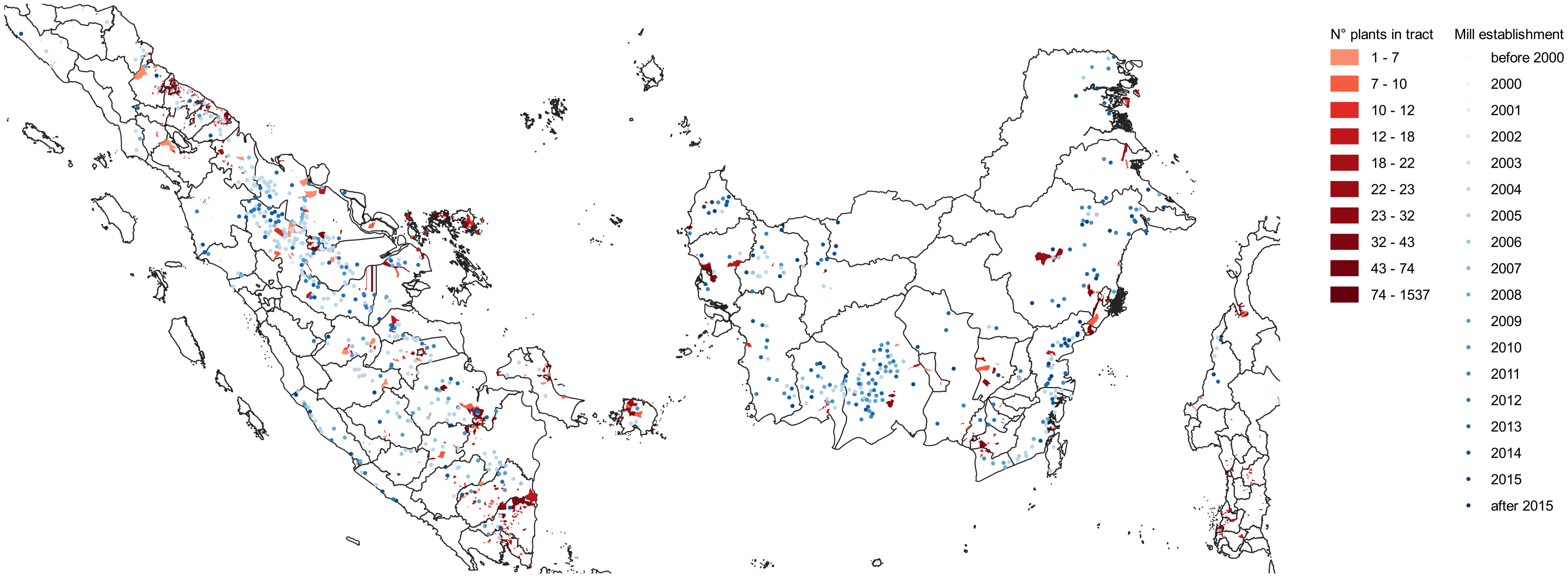
Figures

Figure 1: Construction of the stacked dataset



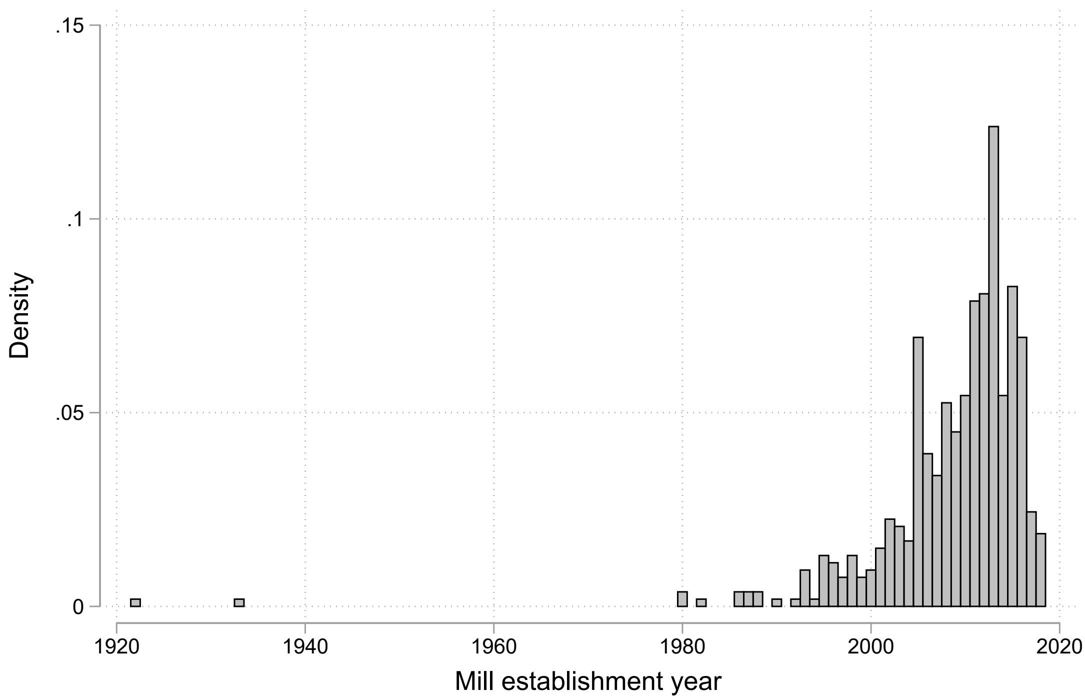
Notes. This figure illustrates the process of selecting manufacturing plant-year observations to act as control observations for individual cohorts. Each year within our study period (2005-2015) defines a cohort of palm oil mills C_c. In the illustrated example, the cohort is C_{2008}. All observations from plants in districts with a new palm oil mill in 2008 form the base of the treatment group in this cohort. All observations from plants based in a district with no new palm oil mill in 2008 form the base of the control group. We then restrict both treatment and control group to observations that fall into the event study window W_{es}, which in our preferred specifications is 5 years before and after treatment. In this example, we thus exclude observations before 2003 and after 2013. In a final step, we remove those observations from the control group according to an exclusion window W_{ex}, that we expect to be on a different trend due to previous treatment. In our preferred specification, W_{ex} covers the same year and the three years before the year of an observation. An observation will not serve in the control group if its unit was exposed to a treatment in the three years before the year of the observation or in that same year. In our example, the 2006, 2007, 2008 and 2009 observations from a control unit are excluded from the control group, because their unit was exposed to a treatment in 2006, i.e., within the exclusion window W_{ex}. Note that we check robustness for an exclusion window that covers the six years before the year of an observation and we also look at a two-sided W_{ex} for three years before and after the year of an observation (see Figure 5).

Figure 2: Palm oil mills (location and establishment year) and number of manufacturing plants in bordering tracts with 1993 district boundaries



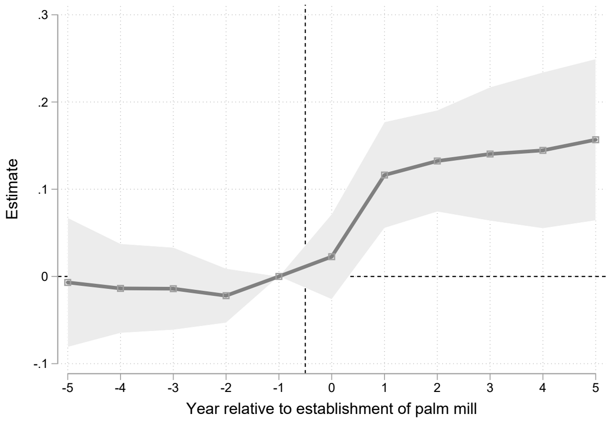
Notes. This figure shows the exact location of palm oil mills and the number of manufacturing plants at the tract (*desa*, village) level in the year 2010, i.e., in the middle of our study period. Darker round dots indicate later establishment dates. We winsorize establishment dates before 2000 into one bin and dates after 2015 into one bin. Lighter colored tracts indicate a smaller number of manufacturing plants and darker tracts indicate a higher number of manufacturing plants, up to over 1500. District borders from 1993 are shown. Treatment is assigned at this level. For our analyses we collapse all other data back into the boundaries.

Figure 3: Distribution of palm oil mill establishment years in the palm oil mill panel

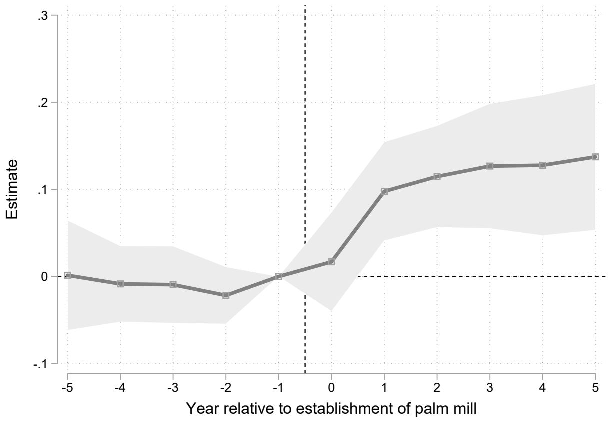


Notes. This figure shows the distribution of establishment years of palm oil mills in our palm oil mill panel. We use establishment dates between 2005 and 2015 that are based on administrative records for the definition of our treatment cohorts. Remaining establishment dates are based on a range of sources, including satellite imagery, journal articles, company reports, mill installation contractor websites, and government websites.

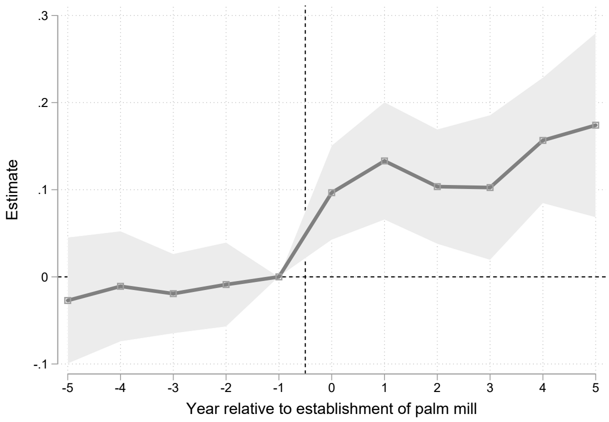
Figure 4: Dynamic effects of palm oil mill establishment on sales, labor productivity and total factor productivity



(a) Sales



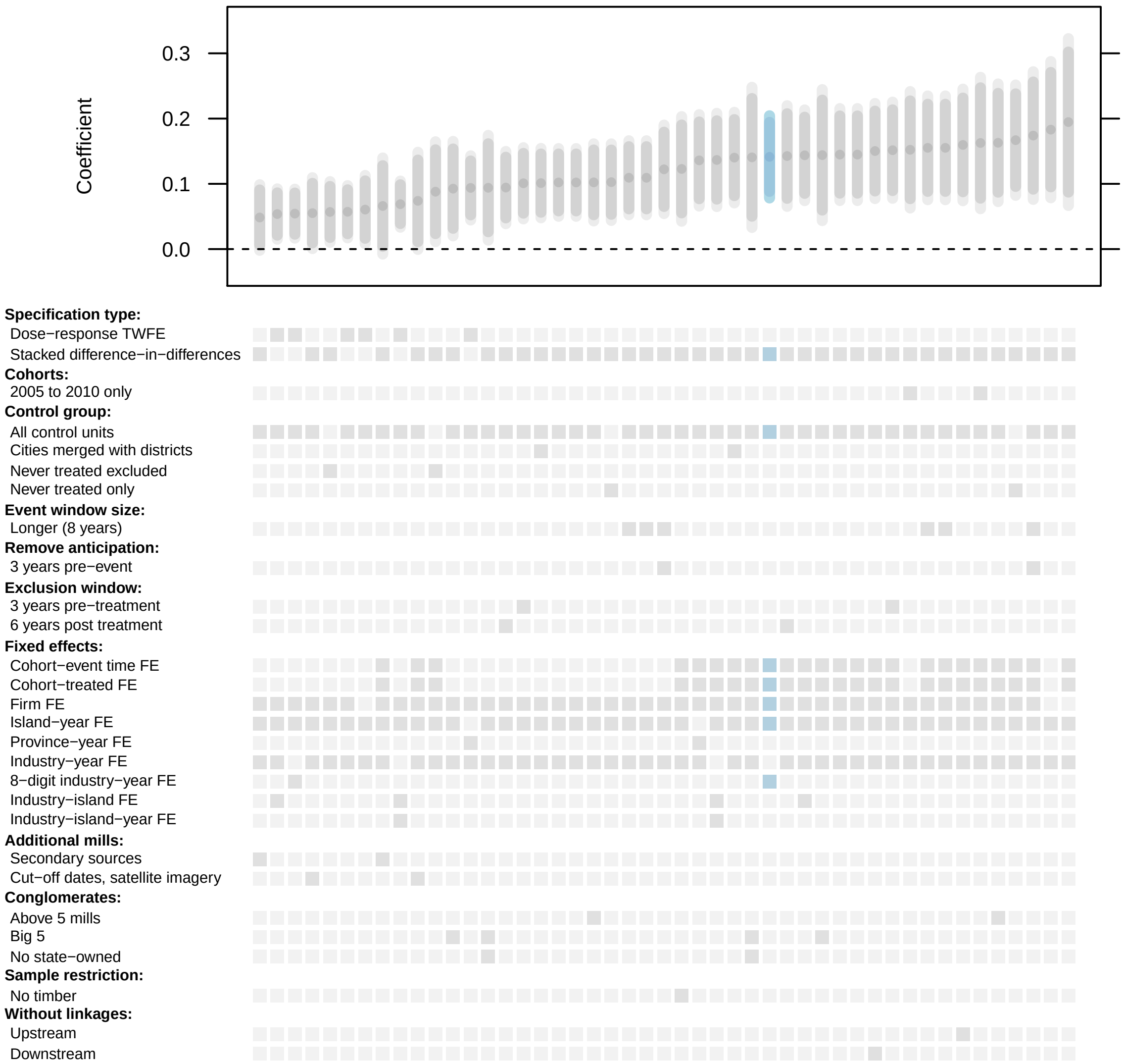
(b) Labor Productivity



(c) Total Factor Productivity

Notes. These figures show the dynamic effects of a new palm oil mill on non-palm oil manufacturing plant performance. The coefficients come from Equation 1 and are also shown in Table 3. Manufacturing plants are grouped into treatment and control groups for each treatment cohorts. Cohorts are stacked relative to event time rather than calendar time. Panel (a) shows effects on the natural log of sales, panel (b) shows effects on the natural log of labor productivity, and panel (c) shows effects on revenue total factor productivity, estimated with the Levinsohn-Petrin estimator, with Ackerberg-Caves-Fraser correction and electricity consumption as the instrument. The year before treatment is used as the base year and the vertical dotted line indicates the timing of the treatment. The 95% confidence band is shown in lighter grey.

Figure 5: Coefficients from different specifications with sales (log)



Notes. This figure shows coefficients from regressions of the natural log of the annual sales of non-palm oil manufacturing plants on a binary treatment indicator for a new palm oil mill in the same district. The 90% confidence interval is marked with a darker bar and the 95% confidence interval is marked with a lighter bar. Our baseline stacked difference-in-differences regression (Equation 2) is marked in blue. We check robustness of this baseline to specification changes along different categories indicated by darker dots in the lower panel. We do this holding all baseline regression features constant and varying only one feature at a time. We show coefficients from regressions (i) on the count (running total) of palm oil mills per district with a dose-response two-way fixed effects specification; (ii) with a sample restricted to cohorts between 2005 and 2010 for balanced pre- and post-periods (our manufacturing plant sample runs up to 2015); (iii) with city districts merged into rural districts; (iv) with never-treated districts excluded (baseline), included, and never-treated only in the control group; (v) with a longer event window size of eight years pre- and post-treatment; (vi) excluding the three years before treatment from the pre-post comparison (we also change the event window to eight years for this robustness check, since otherwise we expect the pre-window to have too little variation); (vii) with different rules for excluding observations from treatment and control groups compared to baseline, i.e., three years exclusion before treatment and six years after treatment; (viii) with different combinations of fixed effect structures, excluding cohort-event time and cohort-treated FE and firm FE, or including FE at the province- rather than island-level, higher resolution industry-year FE, and FE at the industry-island-year level; (ix) with samples including additional sources for establishment, i.e., secondary sources and satellite images in combination with lists of palm oil mills operating at a certain date that we use as a cut-off for the exclusion window; (x) with samples restricted to mills from corporate groups larger than five mills, larger than 25 mills, and larger than 25 mills, but without state-owned company, PTPN III.

For Online Publication

A Additional robustness checks

Figures F.1 and F.2 show our main set of robustness checks for the outcomes, labor productivity and total factor productivity. The labor productivity set is the same as for sales (see Figure 5). For total factor productivity, we add robustness checks on different estimation methods for the residual. Note that the fixed assets variable is missing in a third of our sample and, therefore, some of our robustness checks for total factor productivity are more prone to issues with reduced statistical power than for the other outcomes.

A.1 Standard dose-response fixed effects

We also estimate the dose-response equivalent of our stacked difference-in-differences. For this, we create a running total of mill shocks at the district level. Here treatment variation comes from switches in this running total. Our results are robust to using a dose-response framework, but point estimates become attenuated. Since our treatment does not act as a pure level shifter, but exhibits dynamics over several years, we expect some of this attenuation to stem from control units on a different trend because of previous or anticipated treatment. We could still worry that our stacked research design introduces a selection into control areas that have a lower frequency of mills. Those tend to be either frontier or mature areas, rather than boom areas. In order to rule out the possibility that this type of selection effect is driving the difference between the stacked and the dose-response design, we also change the rules for excluding observations by changing the parameters of the exclusion window. We find that stacked regressions with different exclusion windows lead to similar point estimates as our baseline, particularly when including cohort-treated and cohort-time fixed effects, which are not available in the dose-response design.

A.2 Adding data sources on mill establishment dates

Our main estimation sample is based on establishment dates collected from administrative records (see 3 for a detailed breakdown of sources and sample size). We could be worried that mills for which we have found administrative data are systematically different from other mills and, therefore, our main estimates would be limited in external validity for the whole of the Indonesian palm oil sector. We therefore collected additional establishment dates and ran two robustness checks: (i) we include establishment dates from secondary sources, such as company reports in the estimation sample, (ii) we add establishment dates based on visual inspections of satellite images to the estimation sample and apply our exclusion window based on lists of all mills operating by 1999 and by 2004 respectively, for which we do not have the exact establishment dates. Our estimates remain robust, but standard errors increase and point estimates become attenuated.

A.3 Balanced event window

In our baseline specification, we use all available mill establishment events up to 2015. We only have manufacturing plant data up to the same year. We therefore also check robustness of our results to restricting the sample to cohorts up to 2010, in order to assure balance in terms of time variation between cohorts. Our point estimates are robust to this change.

A.4 Different TFP estimation methods

We estimate revenue-based total factor productivity with standard methods from the production function literature (Ackerberg et al. 2015; Levinsohn and Petrin 2003; Wooldridge 2009), using either electricity consumption (in kWh) or intermediate inputs (value in Rp) in the control function. We exclude total factor productivity based on the Olley and Pakes (1996) method in our main set of outcomes in a pre-analysis step, because of a high number of missing observations and issues with order of magnitude changes due to recordings changing between 1000 Rp and in Rp. We run robustness checks using these different TFP residuals and find that estimates tend to become attenuated when we deviate from our baseline choice, but stay as precise.

B Empirical methods background

B.1 Comparing the stacked and the dose-response fixed effects designs

Our stacked difference-in-differences design has three major advantages: (i) we can exclude units that are still or already on a different trend from acting as controls, (ii) we can investigate pre-trends in a setting with repeated treatment, (iii) we can exercise control over variance in the individual cohorts due to the length of the pre- and post-event window, and thereby reduce differences in regression weights between cohorts. Other applied examples of the stacked design can be found in Cengiz et al. (2019), Deshpande and Li (2019), Fadlon and Nielsen (2015, 2019), Gormley and Matsa (2011), and Jensen (2018).

B.1.1 Regression weights and heterogeneous treatment effects

Our stacked design also helps counter concerns about bias arising from the combination of heterogeneous treatment effects and the weighting mechanics of a standard two-way fixed effects regression. Goodman-Bacon (2018) shows that two-way fixed effects regressions with variation in timing implicitly consist of comparisons between all combinations of early treated, late treated and untreated units.

Regression weights on the sub-effects generated by these comparisons depend on the size of the respective sub-sample, but also on the variance of treatment. In the Goodman-Bacon (2018) case, the treatment variable is a binary indicator. In the staggered research design treatment, variance is therefore driven by how long treatment is turned on in the respective comparison sub-sample. If treatment turns on early or late in the sub-sample window, this results in lower treatment variance and thus a lower weight. Therefore, units that are treated in the middle of the study period have higher weights than those treated at the beginning or at the end. When the timing of treatment adoption is not random, sub-samples at the beginning or the end of the study period may be systematically different from those in the middle. A similar logic applies in the cross-section. Since there is repeated treatment, our treatment variable in a two-way fixed effects regression has to be the count of mills per district. We therefore necessarily work within a dose-response framework or fuzzy difference-in-differences (de Chaisemartin and D'HaultfŒuille 2018).

Heterogeneous treatment drives average estimates of two-way fixed effects regressions, when sub-effect sizes correlate with the treatment variance in sub-comparisons. For instance, there could be selection on gains, meaning that those units with the highest treatment effects get treated first or those with the highest treatment effects might show the biggest resistance to treatment and therefore get treated last. In many empirical contexts, we expect heterogeneous treatment effects that are not randomly distributed over the study period. The Athey and Imbens (2018) formalization of the staggered difference-in-differences design uses random adoption timing as its key assumption and, therefore, seems relevant only to a relatively narrow subset of quasi-experiments.

Since we expect palm plants to build their most promising palm mills first and since pioneer mills are likely to bring pioneer infrastructure and the largest relative spillovers (selection into gains), we have to make sure that these cohorts do not get higher weights than other cohorts. For our stacked research design, we construct cohort sub-samples manually. Thus, we exercise explicit control over the comparisons being made by our regressions. We create more balanced cohorts that are all limited to 5 years before and after treatment. Therefore regression weights are mainly driven by the share of manufacturing plants that are in these cohorts, their cross-sectional variance, and how long they exist before and after the cohort's treatment date. We expect these to be much less correlated with heterogeneous treatment effects than the weights in standard two-way fixed effects regressions. We also also run robustness checks with "fully" balanced treatment cohorts, for which we have manufacturing plant observations both five years before and five years after the treatment year, i.e., 2005 to 2010 (see Section A.3).

C Data cleaning

C.1 Spatial data

Indonesia is divided into four local administrative levels. There currently exist 34 provinces form the first-level subdivision. These provinces are further divided into 416 *kabupaten* (regencies) and 98 *kota* (cities), which form the second-level administration. We use both regencies and cities together as "districts". The 7,071 kemacatan (subdistrict) constitute the third-level subdivision. 81,262 tracts, or villages (Bahasa Indonesia *desa*) form the lowest administrative level.[30]

Due to administrative reforms starting in 1999, (see Section D.2.1) both province and district codes change between survey years. A crosswalk for district and province codes between 1993 and 2014 can be accessed from the World Bank (World Bank Group 2018). We extend this crosswalk until 2016 based on concordances provided by the Indonesia Statistics office (Badan Pusat Statistik 2018). Based on the district crosswalk, we collapse all districts back to their administrative area in 1993 ('base district').

C.2 Treatment: Palm oil mill panel

C.2.1 Data collection

Our main estimation sample is built on administrative records from provincial plantation offices in Indonesia. For robustness checks, we also use data from company reports, satellite imagery, journal articles, mill installation contractor websites, and government websites.

C.2.2 Constructing the stacked data set

We manually create data sets for each cohort of palm oil mill establishments, according to the rules described in 2.1. Afterwards, we append individual cohort data-sets into a pooled data-set. We expand individual manufacturing plant time series by the number of years necessary for our different exclusion rules, i.e., six years before and three years after. As described in 2.1, we want to make sure that observations in the control group are not influenced by an earlier treatment event. We expand plant time series to years before (after) plants appear in (disappear from) the census, so that this rule also fully applies to plants that select into or out of the sample within this time window.

[30]https://www.bps.go.id/website/fileMenu/Perka-BPS-No-90-Tahun-2015.pdf

C.3 Main outcomes: Manufacturing census

Our raw sample consists of 524627 observations between 1993 and 2015 for the sales variable. After cleaning, our base sample contains 492332 plant-year observations. Table E.2 provides an overview of our cleaning steps and details the number of observations dropped in each step. Our estimation sample for regressions with TFPR as the outcome is smaller due to missing observations in the fixed assets variable.

Duplicates IBS imputes values for manufacturing plants that did not fill out the survey from previous years of the same manufacturing plant or from other manufacturing plants that operate in the same 5-digit sector and employ a similar number of workers in the nearest location. Most duplicates between manufacturing plants are from the same survey year and most duplicates for the same manufacturing plant are in subsequent years. Following cleaning methods previously applied to the Indian manufacturing census (Allcott et al. 2016), we drop 4233 exact duplicates based on all variables and 10247 near duplicates based on key variables.[31] We keep one observation per duplicate group if we can clearly identify which one of them is the original, i.e., appears earlier than all other observations in the data.

Redundant questions The manufacturing census questionnaire includes a number of redundant questions. We drop the upper and lower 0.1 percentiles of the ratio of these variables that should report highly similar values. We do this for variables on different types of workers (blue-collar, white-collar, total), for variables on imported, domestic, and total intermediate inputs, and comparing the difference between sales and inputs with the value-added variable. We drop 2412 sales observations in this step.

Workers variable The manufacturing census differentiates between numbers and wages of blue-collar and white-collar workers. In previous work with this data, blue-collar labor has been categorized as "low-skilled" and white-collar labor as "high-skilled". Many of our cleaning routines make use of the number of workers, since this is the most precisely and consistently measured variable in the manufacturing census. We make use of redundant variables measuring the total number of workers by gender, by type of activity and by education level to clean the main workers variable. There are only minor reporting errors in this

[31] These variables are: sales, materials, and workers.

variable. BPS includes only manufacturing plants above 20 workers in the IBS sample. In some years (census years and until the 1990s), manufacturing plants below 20 workers are included in the sample. For our sample period after 1993, there are only two plant-observations with a number of workers below 20.

District codes The manufacturing census data set provides information on each manufacturing plant's province and district code. We use our clean district crosswalk to collapse district codes from all plant years to their 1993 polygon. During our study period, there has been almost a doubling of the number of districts in Indonesia (see Section D.2.1). Also, the statistics agency BPS has changed its district coding system several times during our study period. Many plants, therefore, have outdated district codes in some years. We therefore assign the mode of collapsed 1993 district codes over the whole plant time series and drop 6161 plants that have more than two collapsed district codes that deviate from the mode. There are also 1268 observations for which we could not find a collapsed district code and which we therefore drop from the estimation sample.

Removing palm oil plants Since we are interested in spillovers from palm oil plantations on unrelated manufacturing plants, we remove 8596 plants that produce any palm oil based on our data on all commodities produced by plants.

Fixed assets variable For our TFP estimations we use the estimated value of all fixed assets. Capital variables are substantially less well measured in the Indonesian manufacturing census than other variables. 161291 observations are missing. According to staff at the sub-directory responsible for IBS, there can be unit-of-measurement problems with the capital variables. All IBS variables are collected in 1000 IDR, but some establishments ostensibly have entered numbers that are three orders of magnitude higher or lower.[32]. We drop the lower and upper 0.1 percentile of the fixed asset turnover ratio (sales divided by value of fixed assets) to detect these outliers (22).

Industry codes The Indonesian industry code system, KBLI, is based on ISIC. There have been two main revisions of the KBLI system during the study period: 1997 (basis for KLUI 1997, KBLI 2000 and KBLI 2005) and 2009 (KBLI 2009). These have been adopted for the IBS in 1999 and 2010 respectively. We use concordances from BPS to merge all industry codes into two-digit KBLI 2000 codes, which are

[32]The latter is most likely due to data entry or cleaning mishaps

based on ISIC revision 3. For TFP estimations, we drop 30860 observations that have no industry code or cannot be matched into a unique two-digit industry code.

C.3.1 Total factor productivity

TFP estimates are shown in Table E.3. Our baseline TFP outcome is the residual of a value-added (in Indonesian Rp) production function estimated with the Levinsohn-Petrin (Levinsohn and Petrin 2003) method with Ackerberg-Caves-Frazer (Ackerberg et al. 2015) correction. We use total electricity consumption (sum of electricity from the grid and locally generated electricity in kWh) as the instrument. Because of missing observations in the fixed assets variable and 11227 missing observations for any other variable of the production function, our final sample contains 288932 non-missing observations for the TFP outcome.

D Background information

D.1 Palm oil industry

Elasticity of palm oil demand Palm oil is a substitute for other vegetable oils and even for petroleum diesel in countries that have quotas for biodiesel. Therefore, Indonesia has faced a highly elastic world demand for this vegetable oil.

Labor intensity of palm oil Between 1951 and 1991, the share of labor employed in the mill fell from 17% to 6% (Corley and Tinker 2016).[33] Yields have increased four-fold during the same period, mainly due to the breeding of new varieties, more precise fertilizer and pesticide application, and the introduction of a new pollinating weevil from West Africa (Corley and Tinker 2016; Greathead 1983). The current benchmark is Malaysia, where one worker is needed for every 10 to 12 ha (Byerlee et al. 2016). Palm oil areas have historically seen a steady inflow of migrant workers, initially because of gas and oil extraction and then because of timber concessions. For instance, in Riau on Sumatra, population growth was at an annual rate of 3.4 % between 1990 and 1995 (Baudoin et al. 2019).

Investment needs for a palm mill The typical initial "greenfield" investment in a palm oil business consists of 5-10,000 ha of plantation and a mill, which can typically handle 60 tons of fruit per hour (Byerlee et al. 2016). At maximum capacity, the mill can even handle fruit from an area of 15,000 ha (Cramb and McCarthy 2016). The first harvest is three to five years after planting, and production peaks at 10 years (Corley and Tinker 2016).

Sources of finance in the palm oil sector Typically, financing for palm oil ventures in Indonesia is facilitated by large firms that have access to capital markets(Pramudya et al. 2017). For instance, according to Baudoin et al. (2019), among the 30 % of "grey" companies, for which ownership is unclear, many are backed by money from the big palm oil groups. Another common practice is that palm oil businesses start out with local ownership and are transferred to larger firms as soon as the licensing process has been navigated.

[33] For the impacts of mechanization on plantation work in Malaysia, see Table 11.7 in Corley and Tinker (2016). Until the end of the 1990s, mechanization had reduced labor inputs for transport to mill, weeding, and manuring, but not for harvesting and collection, which went from 76% of labor costs to 93%. Afterwards, most gains in aggregate labor productivity have likely been due to improvements in smallholder practices.

D.2 Institutional background

D.2.1 Pemekaran: the expansion of districts in Indonesia

In the past two decades, Indonesia has gone through an extensive decentralization process that has created a unique natural political economy experiment in terms of the expansion of the land supply, but has also created a number of challenges regarding the harmonization of different administrative maps and codes over the study period.

Following the fall of president Soeharto in 1998, the transitional Habibie administration passed two laws; Law 22/1999 on regional governance, and Law 25/1999 on fiscal relations, granting greater power to the regions. These regulations started the decentralization process. At the beginning of the 2000s, administrative power and financial resources were shifted from the central government to districts. Some of these authorities, especially on land governance, have later been re-centralized.

Districts were also given rights to demand a split-up of their polity. This resulted in a wide proliferation of new local administrative units, known as *pemekaran*.[34] The number of districts (excluding cities) increased by roughly 70% from 242 in 1995 to a total of 416 at the end of 2015.

The numerous splits at the district level can be explained by different drivers. These include political efficiency concerns, financial incentives and bureaucratic rent-seeking (Fitrani et al. 2005), contesting of the rather arbitrarily defined administrative boundaries outside Java (Booth 2011), and the resulting ethnic heterogeneity within districts (Bazzi and Gudgeon 2018; Pierskalla 2016).

D.3 Palm oil licensing

Our identification strategy leverages the phenomenon of large "land banks" in the Indonesian palm oil sector[35]. Large palm oil groups often hold almost as many hectares of land in undeveloped concessions as they operate on the ground. Sometimes, this land is bought from smaller companies that have been created only for the purpose of acquiring and selling a concession. Palm oil group headquarters then build out their portfolio of potential plantations, mostly according to market conditions and local land suitability. Therefore, the timing of mill establishment is exogenous, conditional on fixed effects that capture political, economic, and

[34]See Bazzi and Gudgeon (2018) for a detailed description of this process.

[35]With Article 14 of its New Plantations Law (UU No. 39/2014), the Indonesian government intended to limit this practice by setting a six-year deadline for idle concessions (Consulting 2015)

infrastructure shocks at the regional level.

However, even in the case that a palm oil company in our sample starts developing its plantation immediately after getting a concession, the timing of this start of operation is unlikely to be driven by local factors that could also be driving the performance of unrelated manufacturing plants. This is due to a large number of administrative sources of exogenous delays and regulatory obstacles in the permitting process, and even weather conditions, that can slow down the construction of mills by several months[36] In the following, we provide a short description of idiosyncratic obstacles in the palm oil licensing process and the delays they introduce, based on a detailed description by Paoli et al. (2013).

As described by Burgess et al. (2012), land planning has been decentralized since the fall of Suharto in 1998. In their paper, they describe the cumbersome process of district splitting. The process of licensing a palm oil business is similar, in that it involves sign-off at different levels of governance (national, province, most importantly districts, and villages). The first permit companies have to get is a "Location Permit" (*Ijin Lokasi*) from the district administration. This allows them to start negotiating with local communities for access to the land. In parallel, they have to apply for additional local permits, which also includes an environmental impact assessment, after which they can receive an environmental license (*Ijin Lingkungan* from the local office of the national Ministry of Environment[37]. If companies plan to acquire land that legally belongs to the state forest estate[38], they also need to get the "release" permit (*Surat Pelepasan Kawasan Hutan*) from the Ministry of Environment. While they are in the process of acquiring the final licenses from district authorities, companies apply for the main permit they have to get from the National Land Agency (BPN), the Business Use Permit (*Hak Guna Usaha*, HGU). The data on these HGUs has been the subject of court disputes between NGOs and the government and has not been made fully public, despite courts requiring the government to do so. While HGUs are a core business asset for palm oil companies, which they prominently advertise to investors on their website, for environmental NGOs they are the key indicator of those areas most at risk of being deforested in the future.

[36]See, for instance, the description of Anglo-Eastern's Central Kalimantan mill on their website: https://www.angloeastern.co.uk/about-us/our-business

[37]The other local permits are the plantation business license (*Ijin Usaha Perkebunan*) and the land clearance permit (*Ijin Buka Lahan*)

[38]Even if land has been deforested, it is often still zoned as state forest

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83.6, pp. 2411–2451.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D. O'Connell. 2016. "How Do Electricity Shortages Affect Industry? Evidence from India". *American Economic Review* 106.3, pp. 587–624.

Athey, Susan and Guido Imbens. 2018. "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption". arXiv: `1808.05293 [cs, econ, math, stat]`.

Badan Pusat Statistik. 2018. *BPS Code History*. Last accessed 12 September 2018.

Baudoin, Alice et al. 2019. *Review of the Diversity of Palm Oil Production Systems in Indonesia - Case Study of Two Provinces: Riau and Jambi*. Working Paper 219. Bogor, Indonesia: Center for International Forestry Research (CIFOR).

Bazzi, Samuel and Matthew Gudgeon. 2018. *The Political Boundaries of Ethnic Divisions*. Working Paper 24625. Series: Working Paper Series. National Bureau of Economic Research.

Booth, Anne. 2011. "Splitting, splitting and splitting again: A brief history of the development of regional government in Indonesia since independence". *Bijdragen tot de taal-, land-en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia* 167.1, pp. 31–59.

Burgess, Robin et al. 2012. "The Political Economy of Deforestation in the Tropics". *The Quarterly Journal of Economics* 127.4, pp. 1707–1754.

Byerlee, Derek, Walter P. Falcon, and Rosamond Naylor. 2016. *The Tropical Oil Crop Revolution: Food, Feed, Fuel, and Forests*. Oxford University Press.

Cengiz, Doruk et al. 2019. "The Effect of Minimum Wages on Low-Wage Jobs". *The Quarterly Journal of Economics* 134.3, pp. 1405–1454.

Consulting, Daemeter. 2015. *Indonesia's Evolving Governance Framework for Palm Oil: Implications for a No Deforestation, No Peat Palm Oil Sector*. Bogor, Indonesia: Daemeter Consulting.

Corley, R.H.V. and P.B. Tinker. 2016. *The Oil Palm*. Fifth Edition. Chichester, UK: John Wiley & Sons, Ltd.

Cramb, Rob and John F. McCarthy. 2016. *The Oil Palm Complex: Smallholders, Agribusiness and the State in Indonesia and Malaysia*. Singapore: NUS Press.

De Chaisemartin, Clément and Xavier D'HaultfŒuille. 2018. "Fuzzy Differences-in-Differences". *The Review of Economic Studies* 85.2, pp. 999–1028.

Deshpande, Manasi and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs". *American Economic Journal: Economic Policy* 11.4, pp. 213–248.

Fadlon, Itzik and Torben Heien Nielsen. 2015. *Family Labor Supply Responses to Severe Health Shocks*. w21352. Cambridge, MA: National Bureau of Economic Research, w21352.

— . 2019. "Family Health Behaviors". *American Economic Review* 109.9, pp. 3162–3191.

Fitrani, Fitria, Bert Hofman, and Kai Kaiser. 2005. "Unity in diversity? The creation of new local governments in a decentralising Indonesia". *Bulletin of Indonesian Economic Studies* 41.1, pp. 57–79.

Goodman-Bacon, Andrew. 2018. *Difference-in-Differences with Variation in Treatment Timing*. Working Paper 25018. National Bureau of Economic Research.

Gormley, Todd A. and David A. Matsa. 2011. "Growing Out of Trouble? Corporate Responses to Liability Risk". *The Review of Financial Studies* 24.8, pp. 2781–2821.

Greathead, D. J. 1983. "The Multi-Million Dollar Weevil That Pollinates Oil Palms." *Antenna* 7.3, pp. 105–107.

Jensen, Amalie. 2018. "Loaded but Lonely: Housing and Saving Responses to Spousal Death in Old Age". Unpublished Manuscript.

Levinsohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables". *The Review of Economic Studies* 70.2, pp. 317–341.

Olley, G. Steven and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64.6, pp. 1263–1297.

Paoli, Gary D. et al. 2013. *Oil Palm in Indonesia: Governance, Decision Making and Implications for Sustainable Development*. Jakarta, Indonesia: The Nature Conservancy.

Pierskalla, Jan H. 2016. "Splitting the Difference? The Politics of District Creation in Indonesia". *Comparative Politics* 48.2, pp. 249–268.

Pramudya, Eusebius Pantja, Otto Hospes, and C. J. A. M. Termeer. 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.

Wooldridge, Jeffrey M. 2009. "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables". *Economics Letters* 104.3, pp. 112–114.

World Bank Group. 2018. *INDO-DAPOERs*. Last accessed 12 September 2018.

E Appendix tables

Table E.1: Corporate palm oil groups and their number of mills

Corporate group	Sumatra	Kalimantan	Oth. island	Total
PTPN III	62	10	5	77
SINAR MAS	24	23	1	48
WILMAR	20	15	0	35
ASTRA AGRO LESTARI	11	14	7	32
SIME DARBY	10	15	1	26
SUPERVENTURE	19	3	0	22
ROYAL GOLDEN EAGLE	22	0	0	22
INCASI RAYA (GUNAS)	13	5	0	18
SALIM GROUP	7	9	0	16
MUSIM MAS	10	6	0	16
GAMA PLANTATIONS	12	4	0	16
FIRST RESOURCES	11	4	0	15
BUMITAMA GUNAJAYA AGRO (BGA)	1	13	0	14
DARMEX AGRO	9	5	0	14
MAKIN	6	7	0	13
KUALA LUMPUR KEPONG (KLK)	8	5	0	13
LONSUM	11	1	0	12
TORGANDA	9	0	2	11
SINAR JAYA AGRO INVESTAMA (SJAI)	10	0	0	10
TRIPUTRA AGRO PERSADA	1	9	0	10
SOCFIN	9	0	0	9
CARGILL	3	6	0	9
SUNGAI BUDI	8	0	0	8
EAGLE HIGH PLANTATIONS	0	8	0	8
SAMPOERNA AGRO	5	3	0	8
MAHKOTA	7	0	0	7
GOODHOPE	0	6	1	7
BEST INDUSTRY	0	7	0	7
SIPEF	6	0	0	6
CITRA BORNEO INDAH (CBI)	0	6	0	6
ANGLO-EASTERN PLANTATION (AEP)	5	1	0	6
TELADAN PRIMA	0	6	0	6
BAKRIE	5	1	0	6
GENTING PLANTATIONS	0	5	1	6
DHARMA SATYA NUSANTARA (DSN)	0	6	0	6
PADASA ENAM UTAMA	5	1	0	6
DUTA MARGA	5	1	0	6
SUMBER TANI AGUNG (STA)	5	1	0	6
KENCANA AGRI	1	4	1	6
UNION SAMPOERNA TRIPUTRA PERSADA (USTP)	0	5	0	5
ANJ AGRO	3	1	1	5
WIDYA	2	0	3	5
PASIFIK AGRO SENTOSA (PAS)	3	2	0	5
OTHER	127	112	17	256
UNKNOWN	263	30	12	305
TOTAL	738	360	52	1150

Notes. This table reports the number of palm mills for all corporate groups that are known to own more than 5 mills. We report their mills on the main palm oil islands Sumatra and Kalimantan and those on other islands.

Table E.2: Manufacturing census cleaning

Cleaning step	Dropped	Sample size
Raw IBS sample (1993-2015)		524627
Duplicate observations on all variables	4233	520394
Duplicate observations on main variables	10247	510147
Inconsistencies between redundant variables	2412	507735
Cleaning geographical identifiers	6807	500928
Removing palm oil plants	8596	492332
Missing fixed assets	161291	331041
Trimming 0.1 percentiles of fixed asset turnover rate	22	331019
Industry code missing or ambigious	30860	300159
Any variable of production function missing	11227	288932

Notes. This table reports cleaning steps from the raw manufacturing census to our base sample. The upper panel indicates cleaning steps for the full sample with the main outcome sales. The lower panel indicates cleaning steps for the TFP sample only. Note that our estimation sample is a stacked version of this base sample. That means observations in the estimation sample are those that fall in the event window five years before and after cohort treatment years (2005-2015) and can be used several times, i.e., in several cohorts. Plant-year observations before and after the event-window are used to clean out control observations that may be on a different trend because of previous or later treatment according to an exclusion window.

Table E.3: TFP estimation: Production function coefficients by sub-sector

Industry	ISIC	N	Labor					Capital				
			LP (1)	LP-ACF (2)	WRDG (3)	OP (4)	OP-ACF (5)	LP (1)	LP-ACF (2)	WRDG (3)	OP (4)	OP-ACF (5)
Food products and beverages	15	62761	0.725	1.104	0.755	0.831	0.991	0.156	0.217	0.155	0.191	0.309
			(0.009)	(0.009)	(0.004)	(0.010)	(0.009)	(0.007)	(0.011)	(0.004)	(0.006)	(0.011)
Tobacco products	16	10721	0.756	1.342	0.777	1.153	1.311	0.125	0.190	0.112	0.147	0.345
			(0.033)	(0.024)	(0.014)	(0.029)	(0.023)	(0.016)	(0.029)	(0.014)	(0.036)	(0.027)
Textiles	17	24606	0.685	1.112	0.707	0.910	0.974	0.141	0.151	0.132	0.171	0.334
			(0.014)	(0.014)	(0.006)	(0.011)	(0.000)	(0.009)	(0.016)	(0.006)	(0.010)	(0.000)
Wearing apparel	18	20432	0.890	1.134	0.950	0.998	1.043	0.095	0.095	0.076	0.128	0.224
			(0.011)	(0.040)	(0.006)	(0.007)	(0.010)	(0.012)	(0.033)	(0.008)	(0.013)	(0.002)
Tanning and dressing of leather	19	7091	0.822	1.008	0.854	0.918	0.965	0.124	0.172	0.113	0.099	0.187
			(0.016)	(0.000)	(0.009)	(0.024)	(0.000)	(0.012)	(0.000)	(0.011)	(0.023)	(0.000)
Wood and wood products, except furniture	20	16953	0.759	1.005	0.786	0.931	0.964	0.171	0.296	0.156	0.179	0.323
			(0.013)	(0.016)	(0.007)	(0.013)	(0.000)	(0.007)	(0.003)	(0.008)	(0.009)	(0.000)
Pulp, paper and paper products	21	4527	0.729	1.132	0.790	0.943	1.079	0.156	0.202	0.124	0.145	0.250
			(0.034)	(0.037)	(0.016)	(0.031)	(0.027)	(0.052)	(0.049)	(0.015)	(0.021)	(0.035)
Publishing, printing and reproduction of recorded media	22	6853	0.849	1.196	0.898	1.073	1.164	0.131	0.140	0.100	0.091	0.196
			(0.032)	(0.027)	(0.013)	(0.033)	(0.015)	(0.016)	(0.032)	(0.012)	(0.029)	(0.017)
Coke, refined petroleum products and nuclear fuel	23	456	0.585	1.069	0.583	0.807	0.921	0.196	0.145	0.100	0.101	0.146
			(0.114)	(0.155)	(0.066)	(0.094)	(0.085)	(0.083)	(0.080)	(0.046)	(0.049)	(0.107)
Chemicals and chemical products	24	12130	0.649	1.037	0.664	0.848	0.987	0.203	0.348	0.177	0.198	0.385
			(0.021)	(0.017)	(0.009)	(0.022)	(0.018)	(0.014)	(0.021)	(0.011)	(0.008)	(0.021)
Rubber and plastics products	25	17825	0.650	0.972	0.685	0.791	0.932	0.148	0.177	0.141	0.142	0.258
			(0.013)	(0.018)	(0.007)	(0.022)	(0.020)	(0.012)	(0.021)	(0.008)	(0.012)	(0.024)
Other non-metallic mineral products	26	18441	0.724	1.198	0.724	0.894	1.062	0.151	0.152	0.154	0.192	0.324
			(0.016)	(0.014)	(0.008)	(0.018)	(0.029)	(0.008)	(0.017)	(0.008)	(0.009)	(0.035)
Basic metals	27	2533	0.716	1.080	0.757	0.916	0.999	0.272	0.238	0.244	0.241	0.335
			(0.044)	(0.039)	(0.025)	(0.050)	(0.053)	(0.027)	(0.047)	(0.026)	(0.045)	(0.058)
Fabricated metal products except machinery and equipment	28	10260	0.713	1.108	0.754	0.946	1.002	0.139	0.215	0.137	0.184	0.304
			(0.021)	(0.019)	(0.010)	(0.024)	(0.016)	(0.014)	(0.023)	(0.011)	(0.015)	(0.019)
Machinery and equipment	29	4358	0.828	1.162	0.892	1.008	1.127	0.170	0.231	0.147	0.187	0.265
			(0.042)	(0.042)	(0.017)	(0.037)	(0.045)	(0.023)	(0.039)	(0.018)	(0.023)	(0.049)
Electrical equipment, office machinery, computers	30/31	2651	0.821	1.122	0.888	0.954	1.012	0.145	0.127	0.112	0.165	0.262
			(0.045)	(0.039)	(0.020)	(0.045)	(0.029)	(0.030)	(0.047)	(0.025)	(0.041)	(0.029)
Radio, television and communication equipment	32	1705	0.627	0.971	0.662	0.902	1.011	0.201	0.225	0.217	0.178	0.186
			(0.032)	(0.061)	(0.022)	(0.035)	(0.048)	(0.058)	(0.063)	(0.029)	(0.056)	(0.053)
Medical, precision and optical instruments, watches and clocks	33	630	0.710	1.033	0.747	0.884	0.952	0.173	0.175	0.164	0.127	0.241
			(0.081)	(0.051)	(0.037)	(0.071)	(0.054)	(0.040)	(0.042)	(0.049)	(0.080)	(0.046)
Motor vehicles	34	3104	0.794	1.279	0.864	1.021	1.198	0.110	0.125	0.092	0.180	0.242
			(0.042)	(0.064)	(0.020)	(0.043)	(0.066)	(0.035)	(0.076)	(0.022)	(0.048)	(0.072)
Other transport equipment	35	3307	0.757	1.080	0.828	0.971	1.096	0.216	0.313	0.181	0.139	0.274
			(0.050)	(0.054)	(0.020)	(0.049)	(0.034)	(0.030)	(0.064)	(0.021)	(0.017)	(0.049)
Furniture and n.e.c.	36	23098	0.809	1.059	0.842	0.940	0.995	0.109	0.134	0.099	0.106	0.203
			(0.012)	(0.019)	(0.006)	(0.013)	(0.005)	(0.012)	(0.022)	(0.006)	(0.013)	(0.006)
Recycling	37	415	0.839	1.058	0.991	0.816	0.909	0.051	0.033	0.024	0.052	0.110
			(0.113)	(0.197)	(0.073)	(0.133)	(0.224)	(0.072)	(0.163)	(0.058)	(0.072)	(0.106)

Notes. This table reports TFPR estimates for two-digit industries (2009 KBLI/ISIC codes) in the Indonesian manufacturing census. We use the Levinsohn-Petrin (Levinsohn and Petrin 2003) and Olley-Pakes(Olley and Pakes 1996) estimators, with and without Ackerberg-Caves-Fraser correction (Ackerberg et al. 2015) and the Wooldridge estimator (Wooldridge 2009). Instruments are total electricity consumption (in kWh) for LP and WRDG estimates and new investments in fixed assets (in Rp) for OP estimates.

Table E.4: Effects of palm oil mill establishment on local manufacturing plant turnover

	Entry		Exit	
	(1)	(2)	(3)	(4)
Mill est. (t-5,t-1)	0.021	0.018	−0.012	0.002
	(0.012)	(0.013)	(0.017)	(0.018)
Cohort-event time FE	Y	Y	Y	Y
Cohort-treated FE	Y	Y	Y	Y
Island-year FE		Y		Y
Industry-year FE	Y	Y	Y	Y
District clusters	285	285	285	285
N	1859939	1859939	1859939	1859939

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on firm entry and firm exit in the same district. Both outcomes are binary indicator variables, that are 1 if a firm enters or exits the sample in a given year and zero otherwise. These coefficients are based on our baseline Equation 2.
The unit of observation in this sample is the manufacturing plant. Observations are pooled over the five years before and after a mill is established. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses. We have yearly observations.
In the estimation sample, manufacturing plants are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time rather than calendar time. We exclude observations from cohorts if they come from manufacturing plants that were treated within five years before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample).
All three columns include cohort-event time FE, cohort-treated FE,and industry-year FE (at the five-digit sector level defined by a plant's main output). Columns (2) and (4) also include island-year FE. We do not include firm FE in order to avoid restricting the sample to plants that exist before and after treatment.

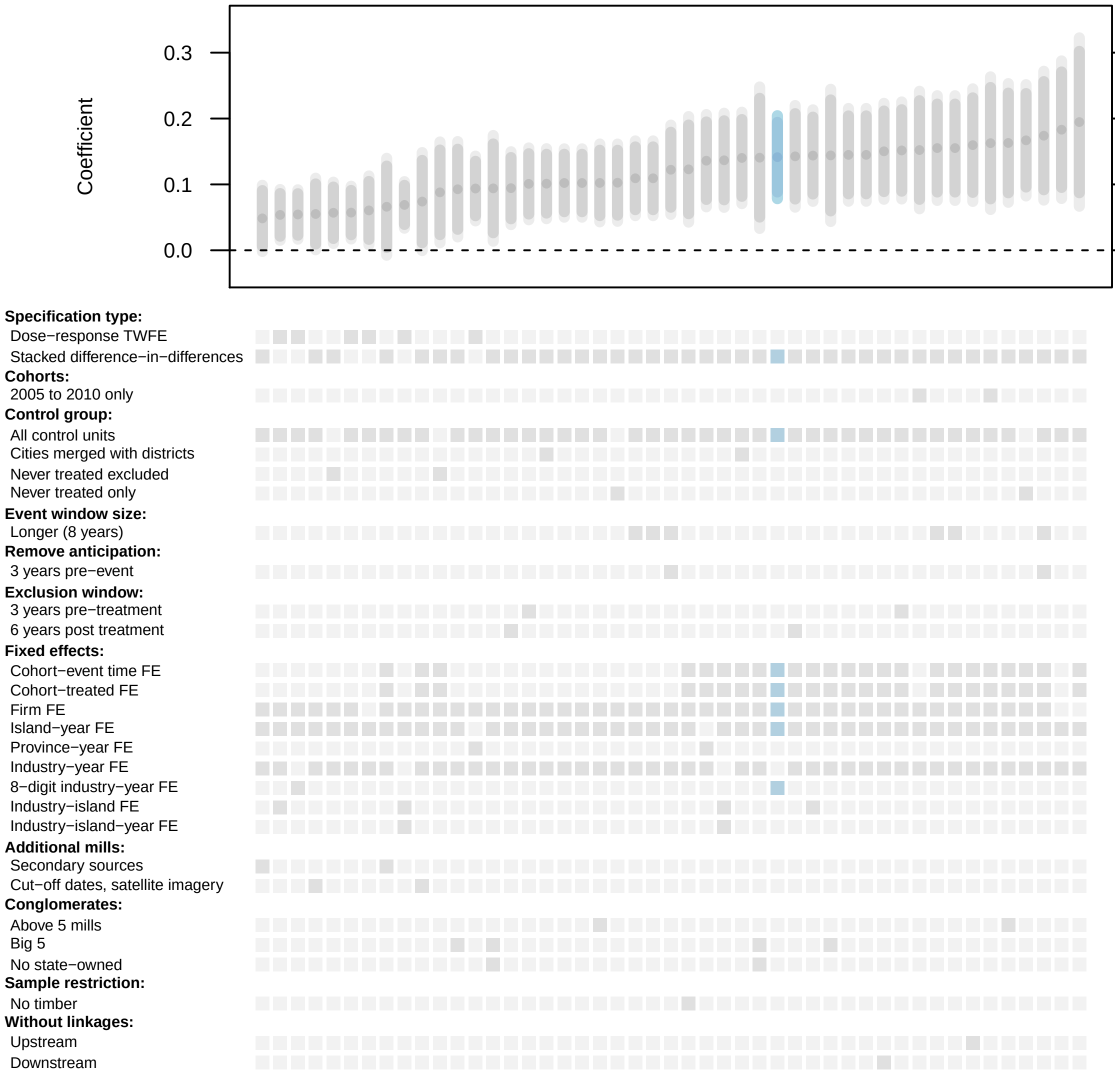
Table E.5: Effects of palm oil mill establishment on downstream and upstream local manufacturing plant performance

	Upstream			Downstream		
	(1) Sales (log)	(2) Labor prod. (log)	(3) TFP (log)	(4) Sales (log)	(5) Labor prod. (log)	(6) TFP (log)
Mill est. (t-5,t-1)	0.092	0.037	0.137	0.096	0.093	0.151
	(0.054)	(0.051)	(0.062)	(0.098)	(0.120)	(0.141)
Firm FE	Y	Y	Y	Y	Y	Y
Island-year FE	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y
District clusters	284	284	282	280	280	276
N	1834709	1834709	1178026	1832441	1832441	1176829

Notes. This table reports the difference-in-differences point estimates of a new palm oil mill on the performance of manufacturing plants upstream (producers of inputs used by palm oil mills) and downstream (buyers of refined palm oil) of palm oil mills in the same district. These coefficients are based on our baseline Equation 2.
The unit of observation in this sample is the manufacturing plant. Observations are pooled over the five years before and after a mill is established. Robust standard errors, adjusted for clustering at the district level, where treatment is assigned, are presented in parentheses. We have yearly observations.
In the estimation sample, manufacturing plants are grouped into treatment and control groups for each treatment cohort. Cohorts are stacked relative to event time rather than calendar time. We exclude observations from cohorts if they come from manufacturing plants that were treated within five years before or three years after the cohort's year (see Section 2.1 for a detailed description of the construction of our baseline sample).
All three columns include our firm FE, island-year FE, and industry-year FE (at the five-digit sector level defined by a plant's main output).

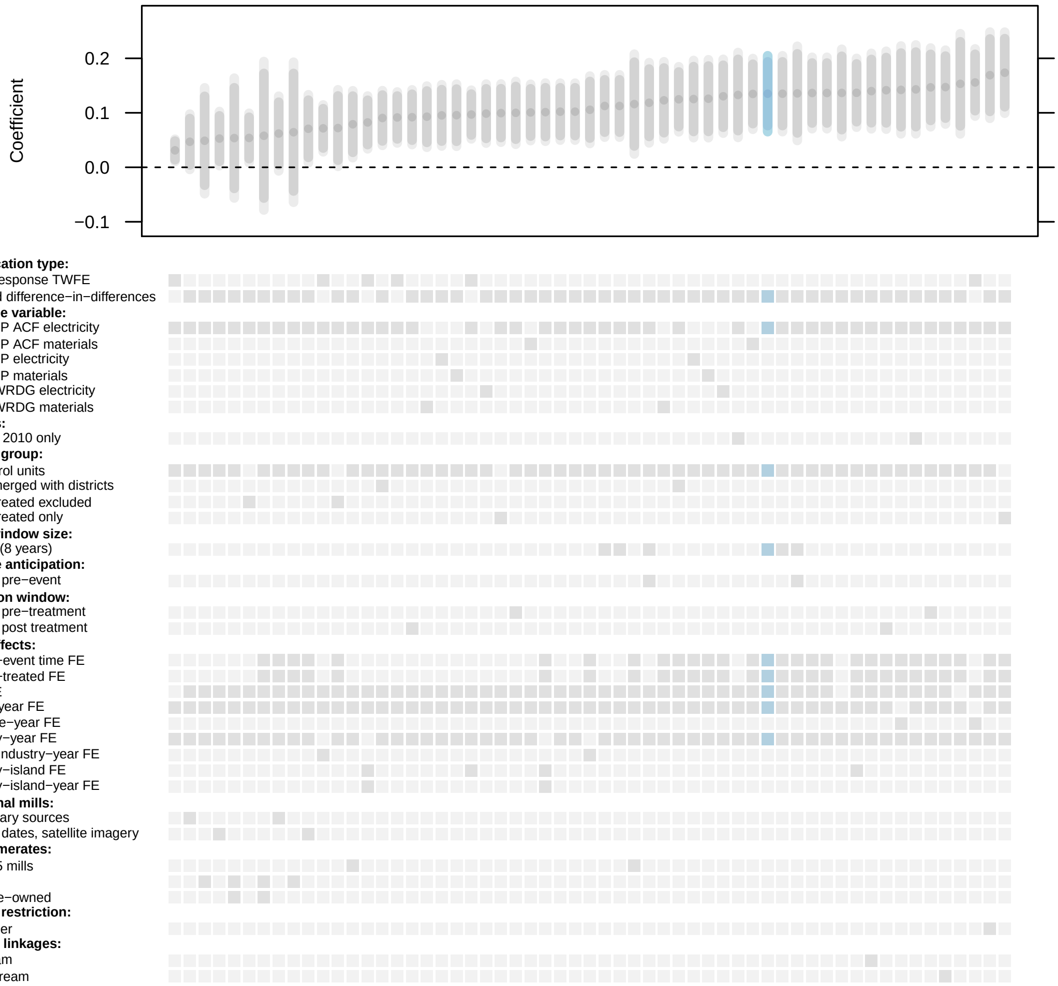
F Appendix figures

Figure F.1: Coefficients from different specifications with labor productivity (log)



Notes. This figure shows coefficients from regressions of the natural log of labor productivity (sales per worker) of non-palm oil manufacturing plants on a binary treatment indicator for a new palm oil mill in the same district. The 90% confidence interval is marked with a darker bar and the 95% confidence interval is marked with a lighter bar. Our baseline stacked difference-in-differences regression (Equation 2) is marked in blue. We check robustness of this baseline to specification changes along different categories indicated by darker dots in the lower panel. We do this holding all baseline regression features constant and varying only one feature at a time. We show coefficients from regressions (i) on the count (running total) of palm oil mills per district with a dose-response two-way fixed effects specification; (ii) with a sample restricted to cohorts between 2005 and 2010 for balanced pre- and post periods (our manufacturing plant sample runs up to 2015); (iii) with city districts merged into rural districts; (iv) with never-treated districts excluded (baseline), included, and never-treated only in the control group; (v) with a longer event window size of eight years pre- and post-treatment; (vi) excluding the three years before treatment from the pre-post comparison (we also change the event window to eight years for this robustness check since, otherwise, we expect the pre-window to have too little variation); (vii) with different rules for excluding observations from treatment and control groups compared to baseline, i.e., three years' exclusion before treatment and six years' after treatment; (viii) with different combinations of fixed effect structures, excluding cohort-event time and cohort-treated FE and firm FE or including FE at the province- rather than island-level, higher resolution industry-year FE, and FE at the industry-island-year level; ; (ix) with samples including additional sources for establishment, i.e., secondary sources and satellite images, in combination with lists of palm oil mills operating at a certain date that we use as a cut-off for the exclusion window; (x) with samples restricted to mills from corporate groups larger than five mills, larger than 25 mills, and larger than 25 mills, but without state-owned company PTPN III.

Figure F.2: Coefficients from different specifications with total factor productivity (log)



Notes. This figure shows coefficients from regressions of the natural log of revenue total factor productivity of non-palm oil manufacturing plants on a binary treatment indicator for a new palm oil mill in the same district. The 90% confidence interval is marked with a darker bar and the 95% confidence interval is marked with a lighter bar. Our baseline stacked difference-in-differences regression (Equation 2) is marked in blue. We check robustness of this baseline to specification changes along different categories indicated by darker dots in the lower panel. We do this holding all baseline regression features constant and varying only one feature at a time. We show coefficients from regressions (i) on the count (running total) of palm oil mills per district with a dose-response two-way fixed effects specification; (ii) using total factor productivity estimates from different methods (LP, ACF, Wooldridge) and with different instruments (materials, electricity) for the two standard sets of FE with and without cohort-specific FE; (iii) with a sample restricted to cohorts between 2005 and 2010 for balanced pre- and post periods (our manufacturing plant sample runs up to 2015); (iv) with city districts merged into rural districts; (v) with never-treated districts excluded (baseline), included, and never-treated only in the control group; (vi) with a longer event window size of eight years pre- and post-treatment (we also change the event window to eight years for this robustness check, since otherwise we expect the pre-window to have too little variation); (vii) excluding the three years before treatment from the pre-post comparison; (viii) with different rules for excluding observations from treatment and control groups compared to baseline, i.e., three years' exclusion before treatment and six years' after treatment; (ix) with different combinations of fixed effect structures, excluding cohort-event time and cohort-treated FE and firm FE or including FE at the province- rather than island-level, higher resolution industry-year FE, and FE at the industry-island-year level; ; (x) with samples including additional sources for establishment, i.e., secondary sources and satellite images, in combination with lists of palm oil mills operating at a certain date, that we use as a cut-off for the exclusion window; (xi) with samples restricted to mills from corporate groups larger than five mills, larger than 25 mills, and larger than 25 mills, but without state-owned company PTPN III.