Building counterfactuals ex-post: Essays in empirical environmental and resource economics

vorgelegt von
M.Sc.
Sebastian Martin Kraus
ORCID: 0000-0003-1161-2988

an der
Fakultät VI – Planen Bauen Umwelt der *Technischen Universität Berlin*
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
– Dr. rer. oec. –

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Johann Köppel
Gutachter: Prof. Dr. Ottmar Edenhofer
Gutachter: Prof. Robert Heilmayr, PhD

Tag der wissenschaftlichen Aussprache: 2. Dezember 2021

Berlin 2021

Abstract

In this dissertation, I make a contribution to the counterfactual-based econometric analysis of panel data. Empirical counterfactuals describe a state of the world without an intervention. In experiments, researchers can create a counterfactual themselves by randomly assigning subjects to a treatment group and a control group for comparison. When researchers can only observe an intervention, they have to construct counterfactuals ex-post. I construct such counterfactuals from observational data for different empirical settings in environmental and resource economics, for which a randomized experiment is not feasible and a control group has to be selected econometrically. I investigate the effects of palm oil on structural change (Chapter 2), of provisional bike lanes on cycling (Chapter 3), of community land titles (Chapter 4) and of palm oil prices on deforestation (Chapter 5), and of air pollution on COVID-19 (Chapter 6).

Causal inference is about finding unbiased counterfactuals. Panel econometrics allows researchers to remove sources of endogeneity with fixed effects. But, except for the simplest research settings, difference-in-differences, researchers have often been unclear about the counterfactual constructed by their panel regressions. Recently, econometricians have started decomposing fixed-effects regressions, explaining them as a weighted average of difference-in-difference estimates. These decompositions have shown that the counterfactuals created in fixed-effects regressions with variation in treatment timing can be biased, notably because previously treated units serve as control units for later treatment cohorts. This is because control units tend to violate the parallel trends assumption, if treatment effects are dynamic, i.e., they do not level off immediately and put units on a different trend for at least several time steps. In this dissertation, I apply a design-based approach to this problem, implemented in a regression framework as "stacked" difference-in-differences (Chapters 2 and 4). This approach allows the researcher to construct "clean" control groups for individual treatment cohorts and thereby mitigate violations of the parallel trends assumption from staggered and repeated treatments. In the Introduction (Chapter 1) of this dissertation, I revisit the difference-in-differences design and its core assumptions. I explain how these assumptions can fail for generalized difference-in-differences settings, which are commonly implemented with standard fixed effects panel regressions. I discuss the stacked difference-in-differences design that allows the researcher to take direct control of the treatment and control group compositions by deconstructing staggered and continuous or repeated treatments into individual treatment cohorts.

Since the "credibility revolution" much of the focus of applied econometrics has been on causal identification. In a Synthesis Chapter 7, I also discuss further challenges to the accuracy of empirical research, beside bias, such as specification searches and limits to generalizability. I propose that, to tackle these challenges, directed acyclical graphs and pre-analysis plans can be important tools for empirical researchers.

Zusammenfassung

Die vorliegende Dissertation ist ein Beitrag zum Feld der kontrafaktischen ökonometrischen Analyse von Paneldaten. Sogenannte Kontrafaktuale (*counterfactuals*) sind hypothetische Beschreibungen eines Zustands der Welt hätte eine zu untersuchende Intervention nicht stattgefunden. In Experimenten können Forschende selbst solche Kontrafaktuale schaffen, indem sie Studienobjekte zufällig entweder einer Behandlungs- bzw. Interventionsgruppe oder einer Kontrollgruppe zuordnen. Wenn Forschende eine Intervention aber nur beobachten können, müssen sie ex-post ein Kontrafaktual aus gemessenen Daten rekonstruieren. In dieser Dissertation schätze ich Kontrafaktuale aus Beobachtungsdaten für verschiedene empirische umwelt- und ressourcenökonomische Kontexte, für die ein randomisiertes Experiment nicht durchführbar ist und eine Kontrollgruppe deshalb ökonometrisch ausgewählt werden muss. Ich untersuche die Auswirkungen von Palmölanbau auf Strukturwandel (Kapitel 2), von provisorischen Fahrradwegen auf den Radverkehr (Kapitel 3), von gemeinschaftlichen Landtiteln (Kapitel 4) und von Palmölpreisen auf tropische Entwaldung (Kapitel 5), und von Luftverschmutzung auf COVID-19 (Kapitel 6).

Mit Methoden der kausalen Inferenz versuchen Forschende, unverzerrte Vergleiche mit Kontrafaktualen anzustellen. Die Panel-Ökonometrie ermöglicht es hier, Endogenitäts-Quellen mit sogenannten festen Effekten (*fixed effects*) zu entfernen. Aber mit Ausnahme des einfachsten Forschungsdesigns aus diesem Bereich, dem Differenz-von-Differenzen-Ansatz, schaffen Analysen mit Panelregression bislang meist keine Transparenz darüber, welches Kontrafaktual sie implizit konstruieren. Neuere methodische Beiträge in der Ökonometrie haben inzwischen gezeigt, dass man Regressionen mit festen Effekten in einen gewichteten Mittelwert einzelner Differenz-von-Differenzen-Schätzungen zerlegen kann. Mit diesen Dekompositionen ist nachgewiesen, dass die Ergebnisse von Regressionen mit festen Effekten verzerrt sein können, wenn der Zeitpunkt einer Intervention für unterschiedliche Kohorten variiert. Das ist insbesondere der Fall, weil in Regressionen mit zweidimensionalen festen Effekten „behandelte“ (*treated*) bzw. zuvor einer Intervention exponierte Einheiten als Kontrolleinheiten für Kohorten mit späteren Interventionen dienen. Diese Kontrolleinheiten können die für dieses Forschungsdesign zentrale Annahme paralleler Trends zwischen behandelten Einheiten und Kontrolleinheiten verletzen, wenn die Effekte einer Intervention dynamisch sind, das heißt nicht sofort abflachen und Einheiten für länger als nur für denselben Zeitschritt auf einen anderen Trend bringen. In dieser Dissertation wende ich einen designbasierten Ansatz auf dieses Problem an, den ich in Regressionsanalysen als „gestapelten“ Differenz-von-Differenzen-Ansatz (*stacked difference-in-differences*) implementiere (Kapitel 2 und 4). Dieses Design ermöglicht es, „saubere“ Kontrollgruppen für einzelne Kohorten zu konstruieren. Dadurch können Verletzungen der Annahme paralleler Trends auch in Kontexten gestaffelter (*staggered*) und wiederholter Interventionen vermieden werden.

Der allgemeine Differenz-von-Differenzen-Ansatz wird in der angewandeten Ökono-

metrie häufig genutzt und meist mit Standard-Panelregressionen mit festen Effekten implementiert. In der Einleitung (Kapitel 1) dieser Dissertation erkläre ich zunächst den einfachen Differenz-von-Differenzen-Ansatz und seine Kernannahmen. Ich erkläre dann, unter welchen Bedingungen diese Annahmen in verallgemeinerten Anwendungen des Differenz-von-Differenzen-Ansatzes (*generalized difference-in-differences*) nicht zutreffen. Anschließend beschreibe ich den gestapelten Differenz-von-Differenzen-Ansatz, der eine direkte Kontrolle über die Zusammensetzung der Interventions- und der Kontrollgruppe ermöglicht, indem gestaffelte und kontinuierliche bzw. wiederholte Interventionen in individuelle Interventionskohorten zerlegt werden.

Seit der „Glaubwürdigkeitsrevolution“ (*credibility revolution*) liegt der Fokus der angewandten Ökonometrie auf kausaler Identifikation. In einem Synthesekapitel 7 diskutiere ich darüber hinausgehende Herausforderungen für die empirische Forschung, wie zum Beispiel fehlende Robustheit von Resultaten durch Spezifikationssuchen und mangelnde Generalisierbarkeit. Ich beschreibe dann Gerichtete Azyklische Graphen (*directed acyclic graphs*, DAGs) und die Präregistrierung von Analyseplänen (*pre-analysis plans*), die hier nützliche neue Werkzeuge sein können.

Contents

Abstract **3**

Zusammenfassung **5**

Contents **7**

1 Introduction **9**
1.1 Motivation 9
1.2 Methodological contribution 13
1.3 References 32

2 Palm oil and structural change: a stacked difference-in-differences approach **35**
2.1 Introduction 35
2.2 Empirical strategy 39
2.3 Data 47
2.4 Results 53
2.5 Discussion 66
2.6 Appendix 69
2.7 References 85

3 An econometric evaluation of cycling infrastructure **89**
3.1 Introduction 89
3.2 Results 91
3.3 Discussion 96
3.4 Methods 98
3.5 SI Appendix 101
3.6 References 114

4 Community land titles and deforestation: stacked difference-in-differences with a control group of eligible areas **117**
4.1 Introduction 117
4.2 Results 119
4.3 Discussion 120
4.4 Materials and methods 123
4.5 SI Appendix 124
4.6 References 132

5 Estimating the deforestation-price elasticity for palm oil with farmgate data 135
5.1 Introduction 135
5.2 Background and theoretical framework 141
5.3 Empirical framework 145
5.4 Data and descriptive statistics 151
5.5 Results 157
5.6 Conclusion 168
5.7 Appendix 171
5.8 References 186

6 Using thermal inversions as a natural experiment to estimate the effect of air pollution on COVID-19 191
6.1 Introduction 191
6.2 Methods 193
6.3 Results 196
6.4 Discussion and conclusion 202
6.5 Appendix 203
6.6 References 206

7 Synthesis and outlook 209
7.1 Common insights 210
7.2 Outlook: new research practices 219
7.3 Conclusion 225
7.4 References 227

List of Tables 230

List of Figures 232

Tools and resources 235
References 239

Statement of contributions 245

List of publications 247

Acknowledgements 249

Chapter 1

Introduction

1.1 Motivation

Humans can think in counterfactuals. Our mind naturally constructs alternative, hypothetical scenarios, for instance when we feel regret, guilt, or relief (Byrne 2016). We also need a counterfactual, when we blame someone for an action or a lack thereof, since that is only justified if we can make a plausible case for a hypothetical, better run of events. Policy-makers want to know the impacts of their interventions – ex ante and ex post. These impacts can only be assessed compared to a counterfactual baseline, sometimes also called a "business as usual". Often, the counterfactual is an implicit assumption, for instance, when a claim is made that a policy "works". In the same vein, applied economists often do not discuss their counterfactual explicitly. In econometrics education, for most of the time in class, we are still taught how to identify regularities in data, rather than how to construct a suitable counterfactual (Angrist and Pischke 2017). This dissertation is about identifying credible counterfactuals. I investigate the limits of common panel regression approaches in econometrics and apply recent methodological advances to construct counterfactuals for five empirical contexts in environmental and resource economics.

What is causation and how can we observe it? Both philosophers and statisticians have proposed that the observation of regularities can teach us about cause and effect. However, detecting regularities is not sufficient to make claims about the causal relationship between two things. We need to know what would have happened to the one thing in the absence of a change in the other thing; we need to know the counterfactual. The computer scientist and philosopher, Judea Pearl notes that already in the eighteenth century, David Hume was struggling with these two elements of causality (2009). Hume's "Treatise of Human Nature" (1739) proposes a theory of causation that relies on the observation of correlation only. He proposes that a flame can be called a *cause* and heat an *effect*, because they can be observed in conjunction, and that repeatedly.[1] Scientists have learned that such correlations do not imply causation. However, the idea that a correlation between two variables means that one is causing the other is intuitive to the researcher's mind and persists in research practice.

An important job for environmental economists is to quantify the costs and benefits of environmental policies. This entails answering questions of the type: What is the direction and the magnitude of the *causal* effect of an intervention D on an outcome Y? When such empirical questions cannot be answered with an experiment, researchers have to

[1] Hume calls this the "constant conjunction" of events.

rely on patterns in observational data. The causal relationships behind these patterns are a priori unclear, since correlations can be due to reverse causality and omitted third factors. Many disciplines, such as epidemiology, have therefore grown sceptical of "causal language" and some authors propose to categorize observational studies as descriptive. Despite this linguistic caution, observational studies often tackle research questions that are essentially *causal* and, therefore, also tend to be interpreted as such in public and even scientific discourse.

In 1748, Hume revised his definition of causality, still building on the old one, but adding, "we may define a cause to be an object followed by another, [. . .] where, if the first object had not been, the second never had existed". In this dissertation, I argue that observational studies should be built on such counterfactuals. That means, that researchers should find a way to describe the world, a firm, or a person in a hypothetical state in the absence of a "treatment". An effect is, then, the difference between such an observed unit in its treated and its counterfactual state. With observational data, such a counterfactual can be constructed based on suitable *control units*. Statistics is then used to obtain a measure of precision for the obtained difference between treated units and counterfactual units, or to judge if we should consider this difference to be systematic or as the result of pure chance. Akin to the evaluation of a randomized controlled trial (RCT),[2] statistics here serves to obtain the standard error of a comparison in means, rather than as a toolbox to fit a function to data. The "credibility revolution" has led this *design-based* approach to become standard in much of applied economics (Angrist and Pischke 2010).

The primacy of research design over estimation in empirical economics is also reflected in the widespread use of the *Rubin Causal Model*, which conceptualizes a treatment effect as a difference in *potential outcomes* under treatment (Y^1) and without treatment (Y^0). Since only one of these potential outcomes can be directly observed, researchers have to construct a counterfactual for the other potential outcome. This is solved by making comparisons at the group level and finding a suitable control group. In this framework, regressions merely provide an estimate for a comparison in (conditional) means between different varieties of treated and control group (Angrist and Pischke 2008). Some newer approaches inspired by common machine learning techniques, such as *matrix completion*, take this philosophy one step further and treat panel data as a matrix with missing potential outcomes that we can impute like missing data (Athey et al. 2018).[3]

The adoption of the *Rubin Causal Model* also changed the language that economists use to discuss their research design: *Dependent variables* are now routinely called *outcomes* and the main *independent variable* is introduced as the *treatment*. The main empirical estimate is referred to as the average treatment effect (ATE).[4] Estimates from regressions on a continuous treatment variable are now often called *dose-response*.

[2]In a randomized controlled trial a researcher picks a treatment group and a control group at random from a population and administers treatment to the treatment group only.

[3]We apply the matrix completion approach as a robustness check in Chapter 3.

[4]Or the average treatment effect on the treated (ATT), depending on the construction of the control group.

Besides the use of the *Rubin Causal Model*, many other new concepts and best practices are mainstreamed into economics from experimental research. This includes the adoption of pre-analysis plans detailing the research design (including cleaning) and providing power calculations (Blair et al. 2018) before data is seen (or generated) and peer review based on pre-analysis plans rather than results. Even the regular seminar question, at which level a researcher should cluster their standard errors, is now being answered referring to the stylized case of a randomized controlled trial: standard errors should get clustered at the level of randomization. Note that, in the absence of randomization, this means that standard errors should get clustered at the level at which the researcher considers a treatment variable exogenous conditional on control variables and fixed effects.[5]

A common element in all of these innovations is the conceptual separation between research design and estimation techniques. Rubin (2008) suggests that, during the design phase of a research project with observational data, all outcome data should be hidden. Further, even if they work with observational data, applied economists now ask themselves what would be the ideal experiment to answer a question, rather than looking for a model that fits their data well. Although this thought experiment helps benchmark research designs, I would argue that, for most applied contexts, it is more useful to ask what a relevant and unbiased counterfactual for treated individuals, firms, regions, or countries before and after treatment would look like. The difference-in-differences model is the formalization of this approach. It compares a change in treated units before and after a treatment with a simultaneous change in control units, i.e., the counterfactual. In order for this counterfactual to be relevant and unbiased, it needs to be built on control units that are (i) representative of the population studied and (ii) exhibit a trend parallel to the trend a treated unit would have followed in the absence of treatment.

A recent literature shows that a variant of this parallel trend assumption needs to hold not only in the stylized context of a difference-in-differences with two groups (treated and control) and two periods (pre-treatment and post-treatment), but also in other panel econometric contexts that can, in fact, be thought of as generalizations of the difference-in-differences design (Abraham and Sun 2018; Callaway and Sant'Anna 2020; de Chaisemartin and D'Haultfoeuille 2020; Goodman-Bacon 2018).[6] This generalization, compared to the standard difference-in-differences setting, can, for instance, consist in including variation in treatment timing and variation in the treatment dose. Variation in treatment timing, also referred to as the staggered adoption of treatment, is routinely discussed as benefiting identification, as it provides additional identifying variation to the cross-sectional variation between treated and control units. Variation in treatment timing can, for instance, be helpful if there is an omitted variable that drives treatment and outcome and has a timing similar to treatment. Consider the case of the simultaneous introduction of a hypothetical tax on deforestation in certain Brazilian municipalities, leaving some municipalities as control units; a standard difference-in-difference setting. One could argue that it is

[5]This is also called "ignorability" in the language of the *Rubin Causal Model*.

[6]See Marcus and Sant'Anna (2020) for a comparison of these different perspectives on the problem and a discussion of earlier related expositions.

likely that a shift in an omitted, unobservable variable, such as a change in political sentiment, could be driving both the adoption of the policy and a relevant outcome, such as deforestation. In a standard difference-in-differences setting, this threat to identification can go undetected, as it would only be visible in a divergence of pre-trends between treatment and control group if the change in political sentiment had built up in the time periods before treatment. If, however, our hypothetical tax on deforestation gets adopted in a staggered manner, for instance by different municipalities in different years, it is less likely that political sentiment, which is principally influenced by national media, would be the driver of policy adoption in each municipality. Further, by restricting the sample to units that eventually get treated, it is more intuitive to argue that control units do not differ systematically from treated units; the only difference between units is when exactly they get treated.

However, Goodman-Bacon (2018) shows that the way this research design is translated into a regression framework can lead to biased estimates. Researchers here typically use a *two-way fixed effects* (TWFE) regression with unit and time fixed effects and a binary treatment indicator with a value of 0 for every time period before treatment and 1 afterwards. This regression generates a weighted average of multiple difference-in-differences between two-group two-period sub-samples, including comparisons that make use of already treated units as control units for units treated later in the study period. If these already treated units are still on a different trend from their treatment, they are not a suitable control unit, since they will create violations of the parallel trends assumption and may even change the direction of the overall regression estimate for the average treatment effect.

Several estimators that can mitigate this problem have been proposed. Baker et al. (2021) provides a summary of these approaches. They typically combine the use of an "event-study" framework that estimates effects in *event time*, rather than *calendar time*. They thereby effectively align treatment timing despite staggered adoption, the estimation and subsequent aggregation of treatment effects for individual cohorts (waves of treatment events), and the restriction of the control group to "clean" units, such as those that do not get treated at all.

In this dissertation, I take a research design- rather than estimation-focused approach. Panel regressions typically deviate from the simple difference-in-differences setup with two groups and two periods. They tend to be based on a mix of control groups, for which it is often unclear if the parallel trends assumption holds. I put the question of what the most appropriate control group for a treated group would be at the center of the design phase. This control group needs to be sufficiently similar to the treatment group that we expect that both groups would have followed a parallel trend in the absence of treatment. In the individual Chapters, I systematically vary control group definitions and expose the stability of the preferred coefficients with *specification charts* (Simonsohn et al. 2020). For Chapters 2 and 4, I also use a new research design, *stacked difference-in-differences*, that allows the construction of credible control groups for each treatment cohort individually.

The dissertation is structured as follows: In the remaining sections of the *Introduction*, I discuss the (generalized) difference-in-differences design and its assumptions in more detail. I then present the stacked difference-in-differences design as a solution to recently identified problems with the unbiasedness of the control group when treatment is staggered or continuous.

The core chapters of this dissertation consist of individual papers that apply these findings to the following empirical questions:

- *Did the Indonesian palm oil boom slow down structural change?* **(Chapter 2)**
- *Can new infrastructure lead to substantial short-term increases in cycling?* **(Chapter 3)**,
- *Did a large-scale community land titling program in Indonesia reduce deforestation?* **(Chapter 4)**,
- *What is the price elasticity of deforestation in the Indonesian palm oil sector?* **(Chapter 5)**
- *Does air pollution worsen COVID-19 outcomes?* **(Chapter 6)**

In a *Synthesis* chapter, I then summarize the results of the papers, revisiting the identification challenges around staggered, repeated treatment in panel regressions.

1.2 Methodological contribution

(Generalized) difference-in-differences: the workhorse of empirical economics

The "credibility revolution" has led to increases in the share of papers published in economics journals and preprint series, such as the NBER working papers, that use empirical methods (Currie et al. 2020). More than 40% of NBER working papers now refer explicitly to experimental or quasi-experimental methods in their text. The (simple) difference-in-differences design is discussed in almost 25% of papers. 60% refer to "fixed effects" methods. Note that some papers describe their research design as based on fixed effects and some refer to the term (generalized) difference-in-differences to describe approaches that are often effectively the same. Other studies describe very similar approaches as "event studies" (7% of all papers) or simply refer to the use of a "distributed lag model". Even when considering that some of these categories are overlapping, a picture emerges of (generalized) difference-in-differences as the most widely used applied research design in economics today; it is the workhorse of the credibility revolution.

In this section, I will briefly introduce the simple difference-in-differences model and the crucial parallel trends assumption. Then I will explain why the parallel trends assumption

needs to hold in many panel econometric contexts, that do not look like a difference-in-differences context at first glance. I will then explain the *stacked difference-in-differences* design that allows a researcher to deconstruct a context with staggered and repeated treatment (or additive changes in the treatment dose) into cohort-specific difference-in-differences. This design allows the definition of criteria to exclude observations from the treatment and the control group to make the parallel trends assumption more plausible a priori. It then allows differences in pre-trends between treatment and control group to be examined visually and in regression form, both for individual cohorts and in a pooled sample.

Difference-in-differences and the parallel trends assumption

The standard two-group, two-period difference-in-differences design is a method for recovering an estimate for a treatment effect, when randomized assignment of a treatment is not possible and it cannot be ruled out that an omitted variable that cannot be measured is driving both the treatment and the outcome. If such an omitted variable was present, $E\left[Y_C^0\right]$, the expected outcome for the untreated controls, would not be a suitable counterfactual for $E\left[Y_T^1\right]$, the expected outcome for treated treatment units. The difference-in-differences research design, instead, builds the following counterfactual to recover an estimate for the treatment effect:

$$E[\delta] = \left(E\left[Y_{T,1}^1\right] - E\left[Y_{T,0}^1\right]\right) - \left(E\left[Y_{C,1}^0\right] - E\left[Y_{C,0}^0\right]\right) \tag{1.1}$$

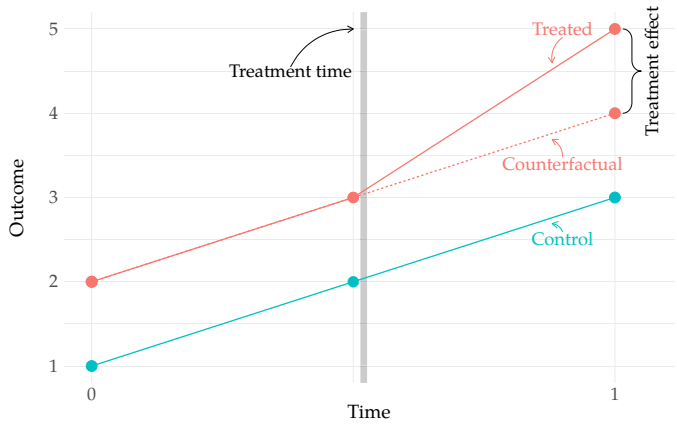
which is based on the double-difference of expected outcomes between treated unit T and control unit C between treatment time 0 and treatment time 1.[7] An estimate for this treatment effect can be obtained by taking the double-differences between the means of treatment and control group before and after treatment (Angrist and Pischke 2008). This research design is widely known in its graphical representation (Figure 1.1).

Note that the research design allows for differences between treatment and control group in baseline outcomes (which in many applications is the case). However, a critical assumption for the double-differences in group means to recover an unbiased estimate of the treatment effect is that the outcome of the treatment group would have followed the same trend as the outcome of the control group in the hypothetical, counterfactual case without any treatment. This is called the parallel trends assumption.[8] This assumption cannot be tested – after all, it is because we do not know the potential outcome $E\left[Y_T^0\right]$, that we are constructing a counterfactual. However, it can be corroborated.

[7]Subscripts indicate treatment time and superscripts indicate the potential outcomes with and without treatment.

[8]See Angrist and Pischke (2008) and Cunningham (2021) for potential outcomes-based textbook treatments of difference-in-differences and the parallel trends assumption.

Figure 1.1: Standard difference-in-differences design



Notes. This figure illustrates the case of a difference-in-differences setup with a single treatment time, i.e., one treatment cohort, and binary treatment, i.e., there is only one treatment group and one control group. The treatment effect is the difference between the treatment value (blue) and the control value (red) post-treatment after subtracting the difference between treatment and control values before treatment. The simplest setup would have four values, i.e., before and after treatment for both treatment and control. However, many studies pool observations before and after treatment and within a control and a treatment group using an average. In our typology, this is still considered a *standard difference-in-differences* design, since there is still only one treatment time and a binary treatment. Here, I show two observations before and one observation after treatment, respectively, for treatment and control group. This allows for the illustration of parallel pre-trends. The dotted line indicates the counterfactual assumed with the parallel trends assumption. This figure is based on code by Baker et al. (2021).

First, if baseline outcomes are different between the treatment group and the control group, researchers should make an argument why this difference is unlikely to be related to factors that could be driving differences in trends (Kahn-Lang and Lang 2020). However, note that balance between the treatment and control group in terms of outcomes or covariates is neither necessary nor sufficient for the parallel trends assumption to hold.

Second, it depends on functional form whether the parallel trends assumption is a priori plausible. Crucially, if mean baseline outcomes in treatment and control group are not the same, the trends cannot hold both expressed in level increases and in growth rates (Kahn-Lang and Lang 2020). For instance, in the case of a hypothetical tax on deforestation in Brazil (Section 1.1), if treated and control municipalities do not have the same deforestation levels, the researcher needs to decide if it is more plausible that both groups of municipalities would have seen the same changes in deforestation in terms of absolute hectares or in terms of changing deforestation rates.

Third, researchers can test, visually and in regression-form, if there are deviations between treatment group and control group before treatment sets in; so-called *pre-trends*. This is typically done by inspecting the evolution of the group means of treatment and control group over time and estimating relative indicators for individual time steps before and after treatment. Note that whether a failure to reject the null hypothesis of no effect for relative time indicators before treatment time t is a meaningful indication for "no pre-trends" critically hinges on statistical power. Therefore, most researchers present "event study plots" that allow for an intuitive visual inspection of pre-trends based on

the evolution of group means or differences in group means and their confidence bands. These plots are now included in most difference-in-differences studies (Currie et al. 2020).

Fourth, even if there are flat pre-trends between treatment and control group, researchers need to argue that there are no omitted variables that drive both treatment and control, whose effect does not show up in outcomes before treatment time t. Note that, in many research settings, this threat to identification is highly plausible. Consider the research question of whether new power plants cause economic growth. It is not unlikely that a third factor, such as a policy or a large investment by a manufacturing business, is driving both the treatment and the outcome. If the start of the policy or the investment is synchronized with the construction of a power plant, this will not show up in the pre-trends of the outcome, economic growth, at least if there are no substantial anticipation effects. One could even argue that there can be a simple case of reverse causality, if economic growth itself is causing the construction of a new power plant. In such cases, researchers can also add high-frequency data to their research design to build counterfactuals based on short-run changes around treatment time t. Such an approach can be seen in Chapter 3, which uses daily data on cycling traffic stretching over years in order to construct both long-run and short-run difference-in-differences.

Two-way fixed effects: a generalized form of difference-in-differences

The difference-in-differences design consists in a comparison of means between a treatment and a control group before and after a treatment. This comparison of group means is typically implemented with a regression of the form:

$$Y_{it} = \alpha_0 + \alpha_1 \text{TREAT}_i + \alpha_2 \text{POST}_t + \beta\,(\text{TREAT}_i \times \text{POST}_t) + \epsilon_{it} \tag{1.2}$$

where y_{it} denotes an outcome for unit i at time t. TREAT_i is 1 for all units i that receive a treatment and 0 otherwise. POST_t marks all observations at time t after treatment. The interaction term $\text{TREAT}_i \times \text{POST}_t$ marks treated units after treatment sets in. Here, α_0 is the intercept of the control group, α_1 captures the difference between treatment and control group before treatment and α_2 captures the slope of the control group. The coefficient of interest β is based on the remaining variation that is based on the remaining difference in the differences.

Angrist and Pischke (2008) write this type of regression in a more generalized form as:

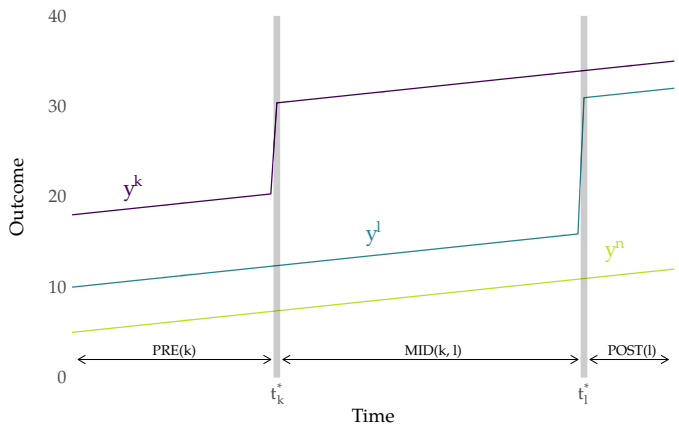
$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{ist} + \epsilon_{ist} \tag{1.3}$$

where Y_{ist} is an outcome for unit i in state s at time t. γ are state fixed effects and λ are time fixed effects (two-way fixed effects). Note that the $TREAT_i$ and $POST_t$ in Equation 1.2 would be subsumed in these fixed effects.

Both regressions can be rewritten in terms of expectations for potential outcomes (Baker et al. 2021) and, assuming linearity, we obtain the unbiased estimate for δ defined in Equation 1.1.

This generalization of the difference-in-differences estimator as a fixed effect regression allows for the inclusion of additional fixed effects, time-varying controls and unit-specific trends. In contrast to the treatment term $\text{TREAT}_i \times \text{POST}_t$ in Equation 1.2 the treatment variable D_{st} in Equation 1.3 can indicate treatment that starts at different times t for different units i or states s and can also have different treatment doses at any time for any unit or state. Around half of a sample of papers published as difference-in-differences had such variation in timing (Goodman-Bacon 2018). However, adding these elements also makes it more complicated to express the estimator in terms of (conditional) differences in means. Figure 1.2 illustrates that it is intuitively unclear, which comparisons are formed in a generalized difference-in-differences setup, even in the simple case of two treatment times.

Figure 1.2: Generalized difference-in-differences with a level-shift treatment

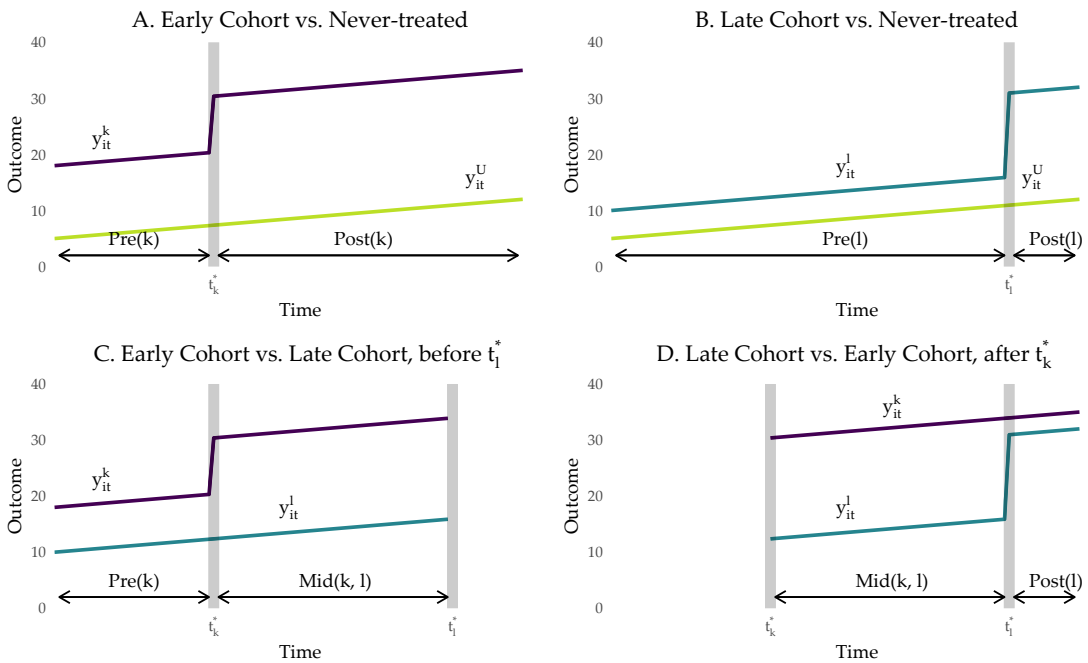


Notes. This figure illustrates a difference-in-differences setup with binary treatment at two times t_k^* and t_l^*, i.e., two treatment cohorts, as illustrated by Goodman-Bacon (2018). For y^k, treatment is at time t_k^*, for y^l treatment is at time t_l^*, and y^n does not get treated. In contrast with Figure 1.1, in this illustration, we have many observations for each y^k, y^l, and y^n in each of the periods $PRE\{k\}$, $MID\{k,l\}$, and $POST\{l\}$. Note that here, the treatment effect is a shift in the level of the outcome, but does not change the trend, i.e., y_k continues to grow at the same rate after t_k^* but from a different level. Whether treatment changes the level or the growth rate of an outcome depends on the study context and on how the outcome is specified. This figure is based on code by Baker et al. (2021).

Goodman-Bacon (2018) has decomposed one of the simpler cases of a generalized difference-in-differences design with binary treatment, variation in treatment adoption, and unit and time-fixed effects. Figure 1.2 illustrates the case with treatment at two times k and l with three types of units. Goodman-Bacon (2018) shows that, for this case, the two-way fixed effects estimator is a weighted average of all possible combinations of two-by-two difference-in-differences in the data (Theorem 1 in Goodman-Bacon (2018)). These combinations include, among others, comparisons between units treated at time t and units treated at time $t - n$. In the case of two different treatment times t_k^* and t_l^* (Baker et al. 2021; Goodman-Bacon 2018) three groups are created: an early treated group treated

at time t_k^*; a later treated group treated at time t_l^*; and a never-treated group, i.e., a group of units that are not treated during the study period. Goodman-Bacon (2018) shows that the regression written in Equation 1.3 forms all four comparisons between these groups, as illustrated in Figure 1.3.

Figure 1.3: Decomposition of generalized (staggered) difference-in-differences with level-shift



Notes. This figure illustrates the Goodman-Bacon (2018) decomposition of the generalized difference-in-differences setup shown in Figure 1.2. It also illustrates a binary treatment at two times t_k^* and t_l^*, i.e., two treatment cohorts. For y^k treatment is at time t_k^*, for y^l treatment is at time t_l^*, and y^n does not get treated. Intuitively, a difference-in-differences is formed around each treatment time t_k^* and t_l^* between each cohort type: early treated y^k, late treated y^l, and never-treated y^n. As in Figure 1.2, and in contrast with Figure 1.1, in this illustration, we have many observations for each y^k, y^l, and y^n in each of the periods $PRE\{k\}$, $MID\{k,l\}$, and $POST\{l\}$. Note that, here, the treatment effect is a shift in the level of the outcome, but does not change the trend; i.e., y_k continues to grow at the same rate after t_k^* but from a different level. Whether treatment changes the level or the growth rate of an outcome depends on the study context and on how the outcome is specified. The figure shows that, for a level shift, treatment effects do not change depending on the comparison group; i.e., it does not make a difference if a treatment cohort is compared to a never-treated or to the other treatment cohort – the treatment effect remains the same. A related but separate issue is that, when treatment effects are heterogeneous, the average treatment effect between sub-effects A, B, C, and D depends on the weights given to each of them. In the illustrated case, the treatment effect on y_l is larger than on y_k. Goodman-Bacon (2018) shows that the weights depend on the size and the variance in the sub-samples of A, B, C, and D. This figure is based on code by Baker et al. (2021).

These four treatment effects enter the overall effect as a weighted average. Their weights are based on the size of the sub-samples and on their variances. This, for instance, means that in the case of a binary treatment with staggered adoption, the sub-effects from the groups treated in the middle of the study period receive higher weights, because the treatment variable is 0 up to the middle of the study period and 1 afterwards, and therefore has the highest variance. Thus, if treatment effects are heterogeneous across cohorts, the average effect based on a two-way fixed effects regression like Equation 1.3 is skewed

compared to an average with equal weights across cohorts.[9] Consider, for instance, the case in which there is a dynamic of selection into gains towards the middle of a program, because, in the beginning, the units with the highest potential treatment effect are still skeptical about enrollment, then towards the peak of a proven program, they enrol and towards the end only those participants with small expected gains remain to enrol.

The strongest effects will be in the middle of the program and, therefore, a researcher using a two-way fixed effects will tend to overestimate the average effect of the program. In this case, the researcher may want to change the weights used to average effects across cohorts, to provide a more relevant average. If effects are highly heterogeneous, researchers may also want to refrain from reporting the average effect all together and present results by cohort.[10]

In addition to this described issue of pooling and weighting sub-effects in the context of heterogeneous treatment effects, and more crucially, there is a potential source of bias that results from comparisons of groups that are on different trends because of previous treatment. Figure 1.3 shows the stylized case, in which treatment only shifts the level of outcome y in treated groups. In the next section, I will discuss what happens if treatment effects are dynamic, i.e., for instance, build up over time or become attenuated over time.

Bias in difference-in-differences with staggered adoption

We have seen that the two-way fixed effects estimator provides a weighted average of all possible two-by-two comparisons (Goodman-Bacon 2018). This includes comparisons between newly treated and already treated units, as you can see in Panel D in Figure 1.3. Remember that, in Figure 1.3, there are no dynamics in the treatment effect; i.e., treatment shifts the level of the outcome y at once and goes back to the same trend as before treatment, just on a different level. In many applications, however, the treatment effect builds up over time or becomes attenuated after an initial "jump".[11] If such dynamics are present, it means that the red line in Figure 1.3 will go on a different trend (or on several different trends) for a number of time steps and only revert back to the pre-treatment trend afterwards, if at all. This means that, in Panel D, the red and green lines will only be on parallel trends after t_l^*, if treatment dynamics are already "wrapped up". If this is not the case, the parallel trends assumption will be violated and estimates will be biased away

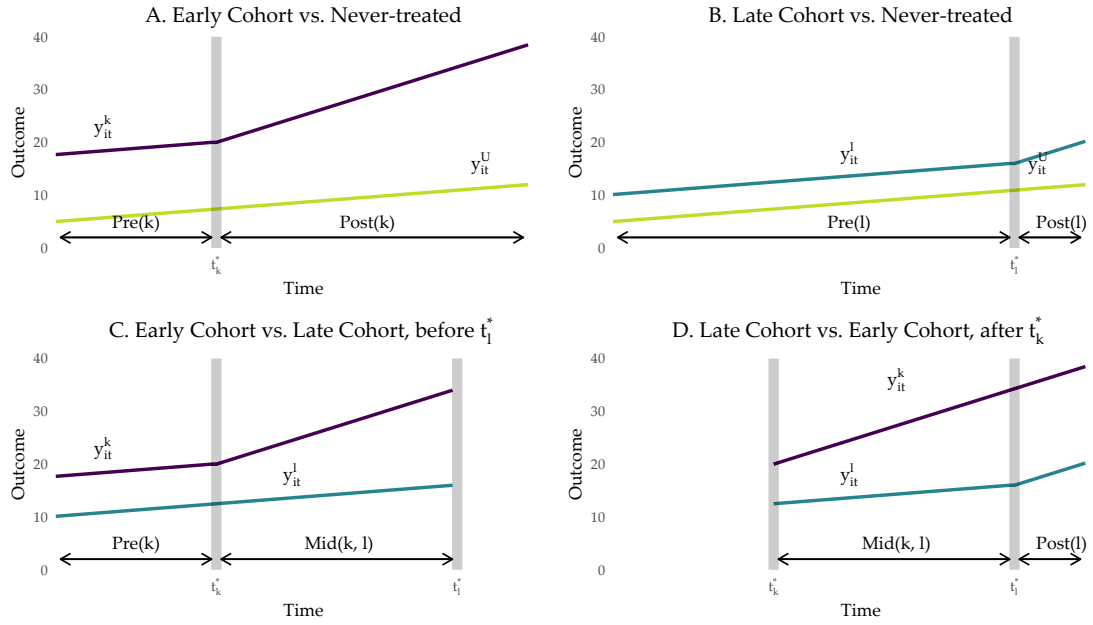
[9] As Baker et al. (2021) notes, panel length alone can therefore shift a coefficient around.

[10] For instance, Cengiz et al. (2019) ("event-by-event analysis" in Appendix D) and Baker et al. (2021) present coefficient plots by cohort.

[11] The literature also refers to these treatment dynamics as "heterogeneous treatment effects in time", as opposed to a level shift, which would be a "constant additive effect". Goodman-Bacon (2018) calls such changes "changes in treatment effect over time". Note that here, authors are referring to changes in the treatment effect of *very same unit* over time and *not* to the separate issue of heterogeneous treatment effects across cohorts, such as when units treated earlier in a study period have higher treatment effects because of selection into treatment.

from the true effect and may even have the wrong sign.[12]

Figure 1.4: Trend-break in staggered difference-in-differences



Notes. This figure illustrates the Goodman-Bacon (2018) decomposition of the generalized difference-in-differences setup, but with a treatment that acts like a trend-break as opposed to the level-shift in Figure 1.2. It also illustrates a binary treatment at two times t_k^* and t_l^*, i.e., two treatment cohorts. For y^k treatment is at time t_k^*; for y^l treatment is at time t_l^*; and y^n does not get treated. Intuitively, a difference-in-differences is formed around each treatment time t_k^* and t_l^* between each cohort type: early treated y^k, late treated y^l, and never-treated y^n. As in Figure 1.2, and in contrast with Figure 1.1, in this illustration we have many observations for each y^k, y^l, and y^n in each of the periods $PRE\{k\}, MID\{k,l\}, andPOST\{l\}$. As opposed to Figure 1.2, in this figure, treatment creates a trend-break, i.e., it changes the growth rate of the outcome, rather than the level only, and therefore, if the outcome is specified in levels, the treatment effect is dynamic. Panel D shows that difference-in-differences with staggered treatment adoption can produce biased estimates when there is such a dynamic treatment effect. This is because the earlier treated y^k acts as a control for the later treated y^k. Since it is still on a different trend from earlier treatment, the parallel trends assumption is violated. The bias resulting from this can be seen by comparing the difference between y^k and y^k before and after treatment. The average gap between y^k and y^k is smaller before than after treatment, but the actual treatment is positive and, thus, the gap should become smaller. The figure also shows that the length of the individual periods is a relevant quantity, if a treatment creates a trend break rather than a shift in levels. This is because difference-in-differences is formed based on an average before and after treatment. This average will become larger over time in the case of a trend-break, i.e., a change in growth rates. This figure is based on code by Baker et al. (2021).

Figure 1.4 illustrates the bias from already treated units acting as controls, if treatment changes the trend in outcome y rather than the level only. Units treated at t_k^* serve as control units for units treated at t_l^*. The two-by-two comparison around treatment time t_l^* is based on the difference in the differences between the two lines before and after t_l^*. Since units treated at t_k^* still trend upwards from this treatment when they serve as controls for units treated at t_l^*, the resulting estimate is biased away from the true effect and the average effect pooled over all two-by-two difference-in-differences may even be turned

[12]If differential pre-trends can be detected in the comparison of a newly treated group with an already treated group due to dynamics from earlier treatment, it may still be the case that the already treated group reverts to the parallel trend "just in time" when treatment for the newly treated group sets in. In this case, the parallel trends assumption may hold despite differential pre-trends. Researcher domain knowledge is needed to judge for how many years units are not suitable as controls because of dynamics from prior treatment.

around. With a suitable control group that has a parallel counterfactual trend to the group treated at t_l^*, for instance from *never treated* units, we would find the true, positive effect.

In the next section, I present a design-based solution that can incorporate the principles behind these estimators: the stacked difference-in-differences design. In Chapters 2 and 4 of this dissertation, I apply the stacked design.

The stacked difference-in-differences design

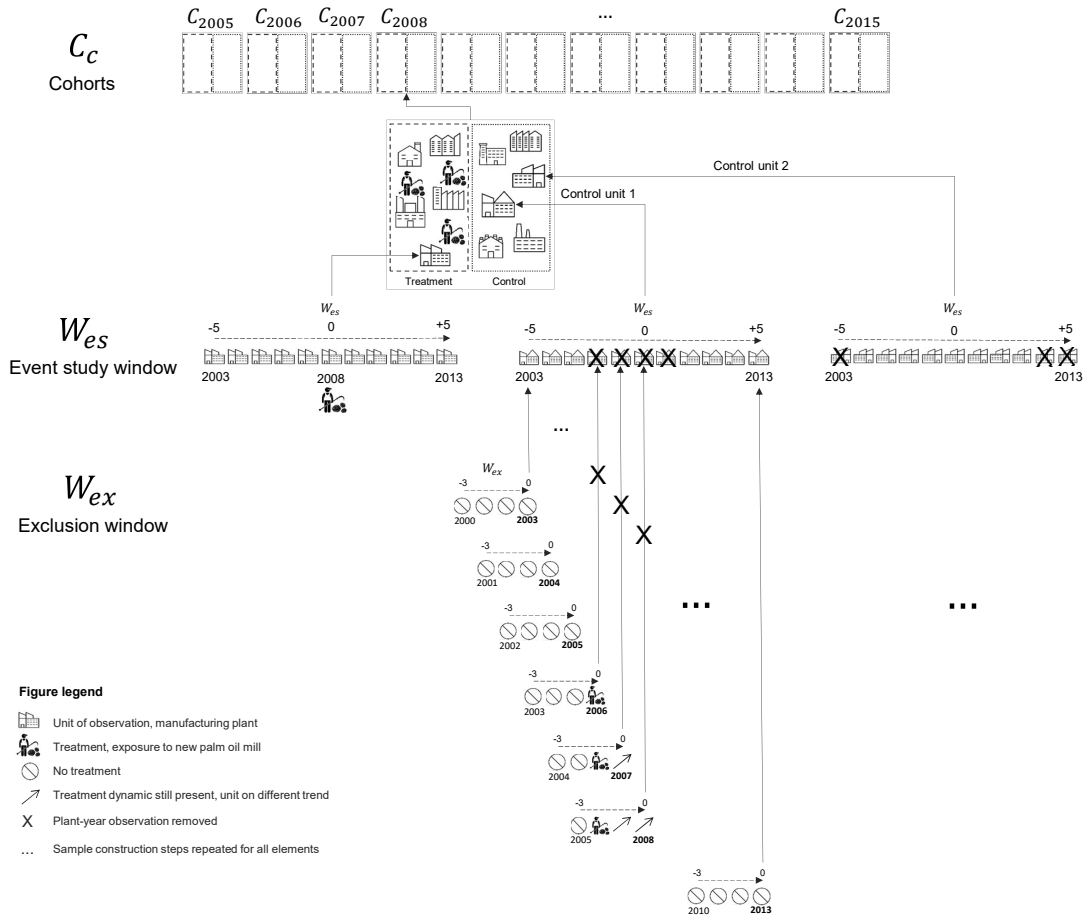
Here, I discuss the *stacked difference-in-differences* design that helps solve the described problems with the two-way-fixed effects estimator. The stacked design allows the researcher to control the composition of the treatment and control group for each cohort. Similarly to some of the estimators proposed in the literature (Abraham and Sun 2018; Callaway and Sant'Anna 2020), this prevents already treated units from re-serving as control units. Crucially, it also allows construction of "clean" treatment and control groups in settings with repeated treatment; for instance, when treatment is continuous and grows in "spurts". This is similar in spirit to the idea of identifying a treatment effect based on "switchers" that change their treatment intensity at time t_s^*, compared to control units that keep their treatment intensity at time t_s^* (de Chaisemartin and D'Haultfoeuille 2018).

Previously, variants of the stacked design have already been used by Cengiz et al. (2019), Deshpande and Li (2019), Fadlon and Nielsen (2015, 2019), Gormley and Matsa (2011), and Jensen (2018). They all create "clean" datasets cohort-by-cohort and then either pool estimates afterwards or stack these datasets in "event time" to run a combined regression. The construction of a stacked estimation sample consists of three steps: (i) selection of suitable treatment and control units for each cohort, (ii) restriction of observations from these units to an *event study window* around a cohort's treatment year, and (iii) the exclusion of a unit's observations as long as it is not back on a parallel trend, according to an *exclusion window*. In this section, I discuss these design choices and their implications.

Constructing the stacked dataset

As mentioned in the previous section, a researcher has to make three key parameter choices when constructing the stacked dataset: (i) how many cohorts to include and from which pool to draw control units, (ii) how many time steps before and after treatment to consider in the *pre-post* comparison, (iii) for how long observations from previously treated units or from units treated in the future should be excluded from the sample. These choices imply important trade-offs in terms of sample selection, sample size balance between cohorts and between the two sides of the event window, and "cleanness" of both the treatment and control group. We illustrate these steps in Figure 1.5, which is set in the research context of Chapter 2.

Figure 1.5: Constructing a stacked sample



Notes. This figure illustrates the construction of a stacked sample of treatment and control groups in the context of treatment in several cohorts, i.e., variation in treatment timing and changes in treatment intensity (or repeated treatment). The figure is from Chapter 2 of this Dissertation. In a stacked sample for each cohort C_c, a clean treatment and control group is constructed. Observations are assigned to the treatment group if they are from a unit that is treated in year c of a cohort; they are assigned to the control group if they are from a unit that is not treated in year c of a cohort. In the next step for both treatment and control groups, only observations that are part of an *event study window* W_{es} around the year of treatment are kept. In this example, the treatment cohort's year is 2008 and observations between 2003 and 2013 are kept. In the third step, for each treatment and control unit observations that are expected to still or already be under treatment dynamics because of prior treatment or anticipated treatment are excluded. For this, an *exclusion window* W_{ex} is defined. In this example, the exclusion window runs three years after treatment. Here, *Control unit 1* is exposed to a prior treatment in 2006 and therefore its observations are removed from the control group in 2006, 2007, 2008, and 2009. *Control unit 2* is treated in 2000 and in 2012 (not fully illustrated). These steps are repeated for each treatment cohort; in this case, from C_{2005} to C_{2015}. Afterwards, the treatment and control groups of these individual cohort datasets are pooled together and pooled regressions like Equation 1.4 are estimates. This stacked design has been proposed by Cengiz et al. (2019), Deshpande and Li (2019), Fadlon and Nielsen (2015, 2019), Gormley and Matsa (2011), and Jensen (2018).

First, the researcher has to set the range, i.e., the study period, and the number of cohorts (Row 1 in Figure 1.5). Each cohort C_c is indexed by time c. For each of the included cohorts, we define a treatment and control group. Observations are assigned to the treatment group, if their units were treated at time c; they are assigned to the control group if their units were not treated at time c. Researchers can decide to restrict the control group to *never treated* units in order to rule out problems with previously treated units acting as

controls. However, if *never treated* units are too different from treated units to serve as credible controls or if not enough *never treated* units are available, researchers can opt to include *later treated* and *earlier treated* units in the control group. In the case of a continuous treatment,[13] the researcher has to decide on a ΔD between two cohort times c, i.e., a change in the treatment variable D between two cohorts C^*_{c1} and C^*_{c2}, from which a unit is considered treated in C^*_{c2} and below which it is considered not treated. For the case of minimum wage increases, Cengiz et al. (2019) this threshold is set at \$0.25; this means that a state is considered treated if the minimum wage is increased by at least \$0.25 in a quarter.

Second, the researcher has to set the size of the *event study window* W_{es}, i.e., the number of time steps before and after the treatment time $c2$ from which observations are included in the comparison (Row 2 in Figure 1.5). W_{es} should be large enough to capture the dynamics of the treatment effect both before and after treatment.

Third, the researcher can define criteria for the exclusion of single *observations* (as opposed to whole *units*) from treatment and control group for which there are reasons to believe that they could still (post) or already (pre) be subject to treatment dynamics, leading to violations of the parallel trends assumption (Row 3 in Figure 1.5).[14] This is the exclusion window W_{ex}, which needs to be applied cohort-by-cohort and observation-by-observation. Other researchers have also excluded all observations from a cohort control group,[15] that are from *units* treated within the event study window W_{es}. This exclusion criterion based on the *unit* is, however, less useful than an exclusion criterion based on individual *observations*. For instance, in the unit-based exclusion mechanism, an observation might be kept in the sample for a cohort C_r, even though it is still subject to treatment dynamics, because it was itself treated in a previous cohort C^*_m just before the beginning of the event study window W_{es} of cohort C^*_r. The same problem would occur if an observation is already subject to treatment dynamics in anticipation of treatment in a later cohort C^*_z. In the observation-based exclusion mechanism, we apply an exclusion window W_{ex} with the same parameter for all individual observations, and it does not matter if previous or later treatment of their unit falls in the W_{es}.

Note that there can be trade-offs between the number of cohorts, the size of the event study window and the stringency of the exclusion window. For instance, researchers have to weigh the number and balance of cohorts with the length of the event study window. The ideal empirical setting would be an intervention rolled out over a limited period of time, with sufficient data before and after the roll-out. This would allow the creation of a fully balanced sample. However, when the temporal extent of the data is limited, there are trade-offs between the length of the study period and the balance of the cohorts studied. In some cases, a researcher may want to evaluate an ongoing policy and recent cohorts need to be included for statistical power. It may also be the case that a policy roll-out has

[13] De Chaisemartin and D'Haultfoeuille (2018) discuss this case as a "fuzzy difference-in-differences".

[14] We do not illustrate removal of observations for control unit 2, which in the example is treated itself in 2002 and 2012.

[15] Observations from treatment units can also be excluded if there is repeated treatment and treatment units are still on a different trend from previous treatment.

started before treatment or outcome get measured. In these cases, specifications with different study windows should be shown.

There can also be a trade-off between the size of the pre- and the post-window in the event study window W_{es}. In many applications, the size of the pre-window should have priority in order to increase confidence in the parallel trends assumption, even if this comes at the expense of missing some dynamics and therefore under- or even over-estimating the treatment effect to some extent.

In many research contexts, the key trade-off in the selection of the number of cohorts, the event study window, and the exclusion window will be between bias and statistical power. If the exclusion window is too stringent, treatment and control groups may become too small to detect an effect. If a research context is well-powered, however, researchers may also opt to fully balance cohorts, i.e., only use treatment and control units in a cohort if none of their observations have to be excluded according to the exclusion window W_{ex}. In this context, researchers also need to take a stance on anticipation effects, since removing them means dropping observations from the event study window W_{es} for some periods before treatment and removing additional observations because they fall in the exclusion window W_{ex}. Researchers should note that dropping observations in these steps can introduce selection bias in the sample. For instance, in the case of repeated treatment, these rules will make the sample more focused on areas with less frequent treatment.

In the stacked design, researchers construct datasets for individual cohorts and append them together, so observations can appear repeatedly in the pooled sample. Cengiz et al. (2019) even estimate state-specific regressions, rather than pooling states that are treated at the same time into cohorts.[16] Researchers cluster standard errors at the cross-sectional level, where treatment is assigned and considered random, conditional on controls and fixed effects, for instance the state, to avoid erroneously increasing the precision of estimates.[17]

Further, in stacked settings in which researchers pool observations from several years into one time t or a single *POST* indicator variable, the canonical recommendation to cluster standard errors in difference-in-differences because of serial correlation (in time) between pooled variables (Bertrand et al. 2004) also holds. An alternative is to collapse all pooled observations into one, for instance by using the mean of all of a unit's observations in the pre- and in the post-group, respectively.

There may also be problems with cross-sectional correlation, for instance, when using data at a fine geographical scale. Researchers can now work with increasingly high-resolution

[16] In their Appendix G, they explain how to aggregate these estimates for a comparison with standard two-way fixed effects results.

[17] In a context of staggered adoption, if conditional on fixed effects and controls, the timing of adoption is random, researchers may also argue that, based on the Abadie et al. (2017) recommendation to cluster standard errors at the level of randomization, standard errors should be clustered at the level of the interaction between cohort C_c and the cross-sectional level where treatment is assigned and considered exogenous; for instance, the state s. However, because of the use of the same observation across cohorts, in the stacked design clustering should still be at the s level.

data, such as GPS-based geographical data. This type of data allows them to make comparisons within small areas, effectively allowing them to control for all unobservable factors beyond these areas with fixed effects. However, depending on the economic phenomenon investigated, at a small geographical scale, spillover effects between units can be substantial and will be amplified when variation between areas is removed. These spillovers can, therefore, violate the *stable unit treatment value assumption (SUTVA)*. SUTVA requires that the change in a treated unit is only driven by its own treatment and not by the treatment of units around it. A related issue is that standard errors can be too small in the presence of patterns of spatial auto-correlation, because the number of observations that actually add precision to the estimate is smaller than in the absence of spatial correlation (Kelly 2020).[18]

Such patterns of spatial correlation will typically not only be a concern for a researcher's standard errors, but will have to be dealt with in the design-stage. This is because it is difficult to define clean treatment and control groups along a smooth gradient of treatment and because the mechanisms that underlie the auto-correlation could be linked to confounders. The stacked research design helps examine the role of spillovers and spatial correlation mechanically. It allows exclusion of units with strong links to other units from the samples for each cohort. I discuss this in more detail in Chapter 7.

Estimation strategy

In its simplest variant, a stacked difference-in-differences regression is specified as follows:

$$y_{istc} = \mu_s \times C_c + \nu_t \times C_c + \eta\,(\text{TREAT}_{sc} \times \text{POST}_{ct}) + v_{ist} \tag{1.4}$$

where i indexes a study unit, s a higher level of geographic (or other) aggregation, such as a state, t indexes calendar time, and c indexes the cohort to which an observation belongs.

y_{istc} is an outcome, C_c is a cohort, TREAT_{sc} indicates treatment, and POST_{ct} indicates that an observation is from a time t after the time t_c^* of treatment of the cohort in which it is serving.

μ_s are state fixed effects and ν_t time fixed effects. Both are absorbed at the cohort level.

The coefficient of interest is η, i.e., the coefficient on the interaction term of the pooled indicators for treatment and time periods post-treatment. Instead of using these pooled indicators, researchers can also estimate the effect per cohort by interacting this term with

[18]See Figure 4 in Kelly (2020) for an illustration of this problem. As Kelly (2020) notes, Tobler's (1970) First Law of Geography says that "everything is related to everything else, but near things are more related than distant things." In contrast, with a fully random data generating process, patterns of correlated noise in a spatial grid look like a "clustered landscape". Intuitively, because of the correlation within clusters, only the areas between these clusters add substantial information to the estimate.

C_c and taking a weighted average of these sub-effects. This allows the researcher to control the weights put on individual cohorts.

Similar to the standard difference-in-differences setting, instead of estimating the coefficient of a single interaction of the treatment indicator TREAT_I with an indicator POST_T for post-treatment observations, we can also include variables that indicate time relative to treatment time t_c^* of a cohort. This allows to plot analyze and plot coefficients over time to examine treatment effect dynamics and check for pre-trends.

Interacting the cohort dummies C_c with all elements of the regression ensures that the analysis yields the same estimate as a manual estimation on individual cohorts and subsequent weighted aggregation. However, in the pooled sample, researchers may also decide to change the fixed effect structure, adding the following elements to the specification:

$$y_{istc} = \mu_s + \nu_t + \text{TREAT}_{sc} \times C_c + \text{POST}_{ct} \times C_c + \theta\,(\text{TREAT}_{sc} \times \text{POST}_{ct}) + v_{ist} \tag{1.5}$$

In this type of specification, unit and time fixed effects are only included at the pooled level and not interacted with the cohort dummy. Intuitively, that means that the researcher does not control (i) for different levels of unobservables at the unit or state level between cohorts, and (ii) for trends in calendar-time specific to the event-window of each cohort.

This specification, instead, includes a fixed effect at the $\text{TREAT}_{sc} \times C_c$ level, that removes differences between the treatment groups and control groups of different cohorts. It also includes a $\text{POST}_{ct} \times C_c$ fixed effect that removes a time trend from pre- to post-event time specific to each cohort.[19]

The coefficient of interest is θ, i.e. the coefficient on the interaction term of the pooled indicators for treatment and time periods post-treatment. Again, the POST_i dummy can also be split up into relative time indicators to analyse the evolution of coefficients over time and to examine pre-trends.

Both the construction of the stacked sample and the selection of fixed effects in regressions create considerable researcher degrees of freedom. Therefore, it seems advisable that researchers show the sensitivity of estimates to these choices and either announce them in a pre-analysis plan or show the robustness of their results to these design and estimation choices.

I apply this stacked research design in Chapters 2 and 4 of this dissertation. All the other Chapters make a case for the unbiasedness of the estimates based on showing robustness to varying the counterfactual. The next section provides an overview of the main chapters

[19]Depending on collinearities, these terms can also be included in specifications like Equation 1.4.

of this dissertation and their methodological contribution to environmental and resource economics.

Objectives and outline

With this dissertation, I make a contribution to the application of counterfactual approaches to the empirics of environmental economics. Some research questions in environmental economics can be answered using randomized controlled trials or quasi-experiments in the narrow sense, such as arbitrary cut-offs for treatment or bugs in assignment algorithms. In many other research settings, researchers have to build credible counterfactuals using non-random variation in treatment between units, including variation in the timing and intensity of treatment. I focus on questions I deem important for environmental policy, and particularly for climate, conservation, and mobility policy, where no natural- or quasi-experiment in the narrow sense is available. I demonstrate how recently discovered limitations of quasi-experimental methods can be addressed. In particular, I show how standard panel regressions can be analyzed from a research design, rather than estimation-based, perspective. I deconstruct panel regressions into a stack of difference-in-differences comparisons and discuss how to understand the parallel trends assumption, treatment exogeneity and the stable unit treatment assumption in these settings. I provide new data points to important debates in environmental and resource economics, for instance on the opportunity cost of tropical forest conservation or the infrastructure elasticity of sustainable mobility behavior.

Ex-post evaluation in environmental and resource economics

Environmental and climate policy-making relies on ex-ante evaluations that weight the costs and the benefits of interventions or describe cost-optimal policy paths to a defined goal. These evaluations are often based on models, such as integrated assessment models for climate policy, that are calibrated with an array of parameters (Nordhaus 1992). While some of these parameters can be based on engineering fundamentals or revealed preferences, for instance from surveys, some other parameters would ideally be estimated based on empirical data and plausibly exogenous variation. This is, for instance, the case for economic parameters such as elasticities of demand and supply, or substitution between factors of production. Environmental economists provide these empirical data points evaluating environmental policy interventions or estimating how certain externalities react to changes in explicit or implicit prices. In 2009, Greenstone and Gayer still described the "improve[ment] of the measurement of the costs and benefits of environmental quality" as the frontier of the field and called for increased use of quasi-experimental methods in environmental economics. Today, more than 30% of environmental papers published in "prominent economics journals" use quasi-experimental methods (Deschenes and Meng 2018). Depending on how the quasi-experimental umbrella is defined, i.e.

only counting truly quasi-random treatment variation or also including variation that is plausibly exogenous based on covariates, this share will be much smaller or much larger. Some environmental economists have the empirical luxury of working with literal natural experiments, such as extreme weather events, pollution exposure due to thermal inversions or wind direction changes. Others can compare units just treated and units just excluded from a program that is rolled out based on a score and an arbitrary threshold (regression discontinuity design). And some researchers also know the exact assignment mechanism of a policy and can, therefore, control for all factors that define selection into treatment. However, researchers working on large-scale environmental policy are also often faced with a dearth of suitable empirical study settings.

For climate policy, in particular, the stakes are high and relevant empirical estimates hard to generate. Not least because of the high stakes of national-level climate-policy with extensive sector coverage, randomized or quasi-experimental settings are rare. Moreover, benchmarked to the ambition described in climate-economic models, climate policy in line with ambitious targets is rare and a "moving target". Many policies have only been implemented in a few, idiosyncratic settings and to a limited extent, raising questions about the external validity of evaluations based on these study contexts. This complicates the task for the empirical environmental economist to find a credible counterfactual, that is both unbiased and has external validity. Consider the debate on the effectiveness of a carbon price. Based on the existing evidence, one *cannot* argue that a carbon tax in the range proposed by climate economists *does not lead* to the promised changes in emissions and investments. Rather, the literature documents certain reductions in emissions already at the low levels and for the limited sectoral coverage (Tol 2021) that have been implemented in the real world. A fundamental problem with a policy proposal that has only been tried out in limited settings is that empirically estimated parameters will have limited external validity. Essentially, carbon prices at levels necessary to meet the Paris climate goals have not been used at scale and, therefore, it is difficult to extrapolate from past experience.[20] Thus, estimating elasticities based on other price shocks can be a more informative approach to gauge how strongly consumption and innovation react to hypothetical carbon prices. I discuss the topic of external validity in more detail in Section 7.1.

In my dissertation chapters, I try to focus on research settings with large, relevant shocks and a credible counterfactual, that teach us something about sectoral climate and sustainability policy. I take a design-based approach to the empirical analysis of observational data and put the parallel trends assumption and a discussion of treatment (timing) exogeneity at the center of my identification argument. I also propose the stacked difference-in-differences design as a remedy against biased control groups due to previous treatment, anticipation or spillovers. This approach allows researchers to examine a wider range of research settings from a design-based perspective, using the analytical tools and heuristics born out of the credibility revolution. It creates transparency on the identifying variation and the critical assumptions needed for a generalized difference-in-differences

[20]This problem is not solved in studies with global panels, such as Rafatya et al. (2020) and Best et al. (2020).

design to be unbiased: parallel trends, SUTVA, and treatment (timing) exogeneity. In the main Chapters 2, 3, 4, 5, and 6 of this dissertation I build the discussion of the identification strategy around these three assumptions. In Chapter 7, I revisit these assumptions and put my findings in the context of related empirical issues in the environmental and resource economics literature.

Research questions and contributions

In the applications of my dissertation, I address empirical gaps in sectoral climate and sustainability policy. Climate policy in the electricity sector is increasingly well understood. Many research gaps remain in other sectors, such as mobility and land use, and on the co-benefits or opportunity costs of climate policy. Empirical evaluations can help understand what works, prioritize interventions, and calculate the net benefits of a policy. I contribute some missing pieces to the empirical sustainability policy landscape by showing that (i) palm oil investments can foster local structural change, (ii) bike lanes can lead to fast, substantial changes in mobility behavior, (iii) involving communities in forest protection with limited use rights is unlikely to be enough to slow down deforestation, and (iv) air pollution has another social cost, in that it worsens Covid-19. I use high-resolution data based on web crawling of cycling counters and ministerial land planning maps, pollution- and deforestation-measurements from satellites, and confidential plant-based input-output data, that allow me to construct credible counterfactuals for each research setting.

With the main chapters of this dissertation, I endeavor to make a contribution to environmental and resource economics by contributing empirical elements to the following broader questions:

Chapter 2 *Does an agricultural boom crowd out manufacturing?* In Chapter 2 we use new data on palm oil mills and the Indonesian manufacturing census to document positive spillovers between a labor-intensive plantation sector and unrelated manufacturing plants, both in terms of sales and productivity. We also use confidential input-output data to rule out that these spillovers only operate through supply-chain linkages. Evaluating the economic impacts of the Indonesian palm oil boom has been hampered by a lack of data. Prior research has used official statistics on palm oil production and planted area (Edwards 2019a), which is representative of the overall distribution of palm oil across Indonesia but less precise temporally, or the cross-section of palm oil mills (Edwards 2019b). We worked with new data on the establishment date of individual *palm oil mills* (Heilmayr et al. 2020), which allowed us to identify the timing of local palm oil booms. Mills are a critical piece of palm oil infrastructure, because palm fruit have to be processed within one day, otherwise they spoil. We exploit clean shocks from the construction of new mills to build transparent counterfactuals for each cohort of new mills. We estimate average

spillover effects on the industrial sector by aggregating effects across cohorts with a stacked difference-in-differences design. The paper demonstrates how to create transparency on assumptions (parallel trends, SUTVA, exogenous treatment timing) similar to a standard difference-in-differences design in a more generalized setting, with a treatment whose intensity varies temporally and in the cross-section.

Chapter 3 *Can new bike infrastructure lead to substantial changes in mobility behavior?* In Chapter 3 we use the roll-out of provisional cycling infrastructure across European cities in 2020 as a policy shock to identify large effects of infrastructure on cycling. The existing literature on the effect of dedicated infrastructure on cycling documents only modest effects (Buehler and Pucher 2012; Dill and Carr 2003). However, researchers have shown that under certain circumstances, such as public transport strikes (Larcom et al. 2017), people change their mobility habits. We confirm this finding in the context of new cycling infrastructure during the Covid-19 pandemic. We crawl daily counts from the websites of municipal cycling counters. We assemble data spanning over a decade. This dataset allows us to construct two counterfactuals: One based on a comparison between treated and untreated cities over the 12 months before and 3 months after treatment; and a second based on both variation in treatment intensity and variation in treatment timing combined with daily cycling rates within the same week. Both counterfactuals confirm that infrastructure can increase cycling substantially and quickly. With our research design, we demonstrate how long panels of high-frequency data can be used to rule out endogeneity.

Chapter 4 *Can we curb deforestation by giving communities some limited access to forest resources?* In Chapter 4, we analyze the early effects of a large-scale community titling program on forest loss in Indonesia. Most of these titles only allow communities to use forests to a limited extent; for instance, to harvest non-timber forest products or run eco-tourism businesses. We show that deforestation rates did not fall after the introduction of titles. We caution that the program in its current design does not provide sufficient incentives for communities to protect forests. We manage to obtain both data on granted titles and maps of areas zoned to be titled in the future. We use the latter category as a counterfactual. The literature on impact evaluation of forest conservation interventions has typically used standard matching techniques to find counterfactuals (Börner et al. 2020). However, these methods cannot control for many time-variant endogeneities that drive selection into treatment, such as local political dynamics. We demonstrate the stacked difference-in-differences approach that allows for flexible integration of controls and fixed effects as an alternative research design.

Chapter 5 *How strongly do palm oil farmers react to changes in prices?* In Chapter 5, we use new microdata on farmgate prices of crude palm oil and palm fruit to estimate the price elasticity of deforestation in the Indonesian palm oil sector. Our new data at the level of the individual palm oil mill, where fresh fruit bunches are processed into crude palm oil,

allows us to run panel regressions and use fixed effects at a high spatial resolution. We document an elasticity of 1.6 of primary forest conversion to palm oil plantations. We find that smallholders and illegal deforestation also react to prices. This indicates that segments of the sector that are difficult for regulators to reach can be incentivized by price-based instruments.

Chapter 6 *Does air pollution drive Covid-19?* In Chapter 6, we use thermal inversions that trap air pollution as a natural experiment to identify the effects of particulate matter on Covid-19 infections and deaths. Many studies and media reports have shown an association between air pollution and Covid-19 outcomes. However, typically they do not account for two first-order sources of endogeneity: First, mobility and economic activity cause both air pollution and infections, creating omitted variable bias. Second, infections lead to people reducing their activities, thereby reducing associated pollution. This creates reverse causality. Any study that relies on fixed effects could, at least partially, account for the omitted variable bias, but could not correct the reverse causality. Therefore, in contrast with the other Chapters of this dissertation, here we use a natural experiment to capture random variation in treatment assignment. Satellite data on pollution and climate models allow us to run a global study at high geographical and temporal resolution. We confirm that air pollution contributes to increases in Covid-19 infections and deaths.

1.3 References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge (2017). *When Should You Adjust Standard Errors for Clustering?* Working Paper 24003. National Bureau of Economic Research. DOI: 10.3386/w24003.

Abraham, Sarah and Liyang Sun (2018). "Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects". *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3158747.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

– (2010). "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics". *Journal of Economic Perspectives* 24.2, pp. 3–30. DOI: 10.1257/jep.24.2.3.

– (2017). "Undergraduate Econometrics Instruction: Through Our Classes, Darkly". *Journal of Economic Perspectives* 31.2, pp. 125–144. DOI: 10.1257/jep.31.2.125.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi (2018). *Matrix Completion Methods for Causal Panel Data Models*. Working Paper 25132. National Bureau of Economic Research. DOI: 10.3386/w25132.

Baker, Andrew, David F. Larcker, and Charles C. Y. Wang (2021). *How Much Should We Trust Staggered Difference-In-Differences Estimates?* SSRN Scholarly Paper ID 3794018. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3794018.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics* 119.1, pp. 249–275. DOI: 10.1162/00335530472839588.

Best, Rohan, Paul J. Burke, and Frank Jotzo (2020). "Carbon Pricing Efficacy: Cross-Country Evidence". *Environmental and Resource Economics* 77.1, pp. 69–94. DOI: 10.1007/s10640-020-00436-x.

Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys (2018). *Declaring and Diagnosing Research Designs*. Working Paper.

Börner, Jan, Dario Schulz, Sven Wunder, and Alexander Pfaff (2020). "The Effectiveness of Forest Conservation Policies and Programs". *Annual Review of Resource Economics* 12.1, pp. 45–64. DOI: 10.1146/annurev-resource-110119-025703.

Buehler, Ralph and John Pucher (2012). "Cycling to Work in 90 Large American Cities: New Evidence on the Role of Bike Paths and Lanes". *Transportation* 39.2, pp. 409–432. DOI: 10.1007/s11116-011-9355-8.

Byrne, Ruth M.J. (2016). "Counterfactual Thought". *Annual Review of Psychology* 67.1, pp. 135–157. DOI: 10.1146/annurev-psych-122414-033249.

Callaway, Brantly and Pedro H. C. Sant'Anna (2020). *Difference-in-Differences with Multiple Time Periods*. Preprint.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs". *The Quarterly Journal of Economics* 134.3, pp. 1405–1454. DOI: 10.1093/qje/qjz014.

Cunningham, Scott (2021). *Causal Inference: The Mixtape*. Yale University Press.

Currie, Janet, Henrik Kleven, and Esmée Zwiers (2020). "Technology and Big Data Are Changing Economics: Mining Text to Track Methods". *AEA Papers and Proceedings* 110, pp. 42–48. DOI: 10.1257/pandp.20201058.

De Chaisemartin, Clément and Xavier D'Haultfoeuille (2018). "Fuzzy Differences-in-Differences". *The Review of Economic Studies* 85.2, pp. 999–1028. DOI: 10.1093/restud/rdx049.

– (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". *American Economic Review* 110.9, pp. 2964–2996. DOI: 10.1257/aer.20181169.

Deschenes, Olivier and Kyle C. Meng (2018). "Chapter 7 - Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges". *Handbook of Environmental Economics*. Ed. by Partha Dasgupta, Subhrendu K. Pattanayak, and V. Kerry Smith. Vol. 4. Handbook of Environmental Economics. Elsevier, pp. 285–332. DOI: 10.1016/bs.hesenv.2018.08.001.

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. DOI: 10.1257/pol.20180076.

Dill, Jennifer and Theresa Carr (2003). "Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them". *Transportation Research Record* 1828.1, pp. 116–123. DOI: 10.3141/1828-14.

Edwards, Ryan B. (2019a). "Export Agriculture and Rural Poverty: Evidence from Indonesian Palm Oil". Mimeo.

– (2019b). "Spillovers from Agricultural Processing". Mimeo.

Fadlon, Itzik and Torben Heien Nielsen (2015). *Family Labor Supply Responses to Severe Health Shocks*. Working Paper 21352. National Bureau of Economic Research. DOI: 10.3386/w21352.

– (2019). "Family Health Behaviors". *American Economic Review* 109.9, pp. 3162–3191. DOI: 10.1257/aer.20171993.

Goodman-Bacon, Andrew (2018). *Difference-in-Differences with Variation in Treatment Timing*. Working Paper 25018. National Bureau of Economic Research. DOI: 10.3386/w25018.

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. DOI: 10.1093/rfs/hhr011.

Greenstone, Michael and Ted Gayer (2009). "Quasi-Experimental and Experimental Approaches to Environmental Economics". *Journal of Environmental Economics and Management* 57.1, pp. 21–44.

Heilmayr, Robert, Kimberly M. Carlson, and Jason Jon Benedict (2020). "Deforestation Spillovers from Oil Palm Sustainability Certification". *Environmental Research Letters*. DOI: 10.1088/1748-9326/ab7f0c.

Hume, David (1739–1740). *A Treatise of Human Nature*. Vol. 1. London.

– (1748). *An Enquiry Concerning Human Understanding*. London.

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

Kahn-Lang, Ariella and Kevin Lang (2020). "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications". *Journal of Business & Economic Statistics* 38.3, pp. 613–620.

Kelly, Morgan (2020). "Understanding Persistence". Mimeo.

Larcom, Shaun, Ferdinand Rauch, and Tim Willems (2017). "The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network". *The Quarterly Journal of Economics* 132.4, pp. 2019–2055. DOI: 10.1093/qje/qjx020.

Marcus, Michelle and Pedro H. C. Sant'Anna (2020). "The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics". *Journal of the Association of Environmental and Resource Economists*. DOI: 10.1086/711509.

Nordhaus, William D. (1992). *The 'DICE' Model: Background and Structure of a Dynamic Integrated Climate-Economy Model of the Economics of Global Warming*. 1009. Cowles Foundation for Research in Economics, Yale University.

Pearl, Judea (2009). *Causality*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511803161.

Rafatya, Ryan, Geoffroy Dolphin, and Felix Pretis (2020). "Carbon Pricing and the Elasticity of CO_2 Emissions". *Institute for New Economic Thinking Working Paper Series*, pp. 1–84. DOI: 10.36687/inetwp140.

Rubin, Donald B. (2008). "For Objective Causal Inference, Design Trumps Analysis". *Annals of Applied Statistics* 2.3, pp. 808–840. DOI: 10.1214/08-AOAS187.

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson (2020). "Specification Curve Analysis". *Nature Human Behaviour* 4.11 (11), pp. 1208–1214. DOI: 10.1038/s41562-020-0912-z.

Tol, Richard S. J. (2021). *Ex-Post Evaluation of Climate Policy*. RePEc Biblio topic.

Chapter 2

Spillovers to manufacturing plants from multi-million dollar plantations: evidence from the Indonesian palm oil boom[†,‡]

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 that this effect is driven by supply chain linkages. Our results are robust in a sample of large corporate groups that assign treatment more independently of changes in local conditions.

2.1 Introduction

A positive shift of comparative advantage in labor-intensive agriculture can crowd out industrial growth as workers relocate to plantations. 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 plants typically make greenfield investments of US$100 million to set up a mill and its adjacent plantations where land is naturally suitable for growing oil palms. We show that these shocks have, on average, led to positive shifts in sales (15%), labor productivity (13%) and total factor productivity (13%) in non-palm oil manufacturing plants.

In 2000, Indonesia still exported more petroleum oil, electronics, garments, and wood products than palm oil. In 2015, palm oil was its largest export with a share of 11% (up from 2% in 2000). In contrast to the Green Revolution, Indonesia's quadrupling of palm oil production since 2000 has not been driven by technological advances, but by an expansion

[†]Sebastian Kraus, Robert Heilmayr, and Nicolas Koch (2020). "Spillovers to Manufacturing Plants from Multi-Million Dollar Plantations: Evidence from the Indonesian Palm Oil Boom". Revise and Resubmit *Journal of the Association of Environmental and Resource Economists* on 23 August 2021.

[‡]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.

of the land supply, often at the detriment of natural forest areas. Indonesia's political and fiscal decentralization process has sped up increases in the land area under palm oil and timber concessions (Burgess et al. 2012).

The land-fueled palm oil boom in Indonesia can be compared to the expansion of soy in Brazil. There, new soy farms have benefited from cheap land at the deforestation frontier in the Amazon. Brazil used to export more cars and aircraft than soy, but the soy share in its exports has since tripled. At the same time, employment in the soy sector has halved due to the introduction of new technologies (GMOs and automation). Bustos et al. (2016) have shown that this release of labor has led to increased employment in the manufacturing sector.

Meanwhile, in Indonesia, palm oil labor intensity on industrial plantations has decreased at a much slower pace. The harvesting and collection of fresh fruit bunches are still largely done 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 by 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. The sector today employs two million people, the large majority of them as plantation workers.

This paper examines how manufacturing plants have reacted to local waves of palm oil expansions in Indonesia. 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

[1]See Appendix 2.6 for more detail on innovations and investments in the palm oil sector that impact factor productivity and substitution elasticity.

[2]See Appendix 2.6 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.

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 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 different research domains in development, resource, and agricultural economics.

First, we provide an empirical test for predictions made in the structural change literature for the case of an expansion of the land supply and an agricultural sector with a strong complementarity between land and labor. 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 for the links between agricultural expansion and industrial growth. Conceptually the case of palm oil is similar to the double cropping of maize in the Bustos et al. (2016) analysis of structural change in Brazil, based on a model in which land and labor are strong complements. In that setting, the labor-saving introduction of genetically modified soy led to an expansion of manufacturing, while the land-augmenting introduction of a second harvesting season for maize led to a smaller manufacturing sector. In Indonesia, land conversion to palm oil is the main driver of the sector's growth. Therefore, the palm oil expansion is comparable to land-augmenting change in Brazil. Like Bustos et al. (2016) we find an overall shift towards employment in agriculture. However, we show that the related increases in blue-collar wages do not need to lead to a contraction of the manufacturing sector. On the contrary, we find increases in sales in 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 that we can measure as 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.

Second, we investigate a slow-down in local structural change as one potential channel of an aggregate resource curse. While the palm oil industry relies on a decentralized network of mills because palm fruit should be processed within a day, local up- and downstream supply chain linkages 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. Researchers have turned to micro-data to investigate the resource curse with spatially explicit micro-data.[4] Allcott and Keniston (2018) use the US census of manufacturers to show that manufacturing is not crowded out

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

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 contexts.

Third, we build on a growing literature examining the socioeconomic effects of the Indonesian palm oil boom. Oil palm plantations have lifted 1.3 million out of poverty (Edwards 2019a), are linked in the cross section 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). We contribute to this literature by investigating the dynamic behavior 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. Our analysis indicates that, in addition to the previously studied immediate local benefits of the sector, the palm oil boom has had important, positive structural impacts on manufacturing output, productivity, wages and tax revenues.

2.2 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 2.6 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.

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 2.6 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.

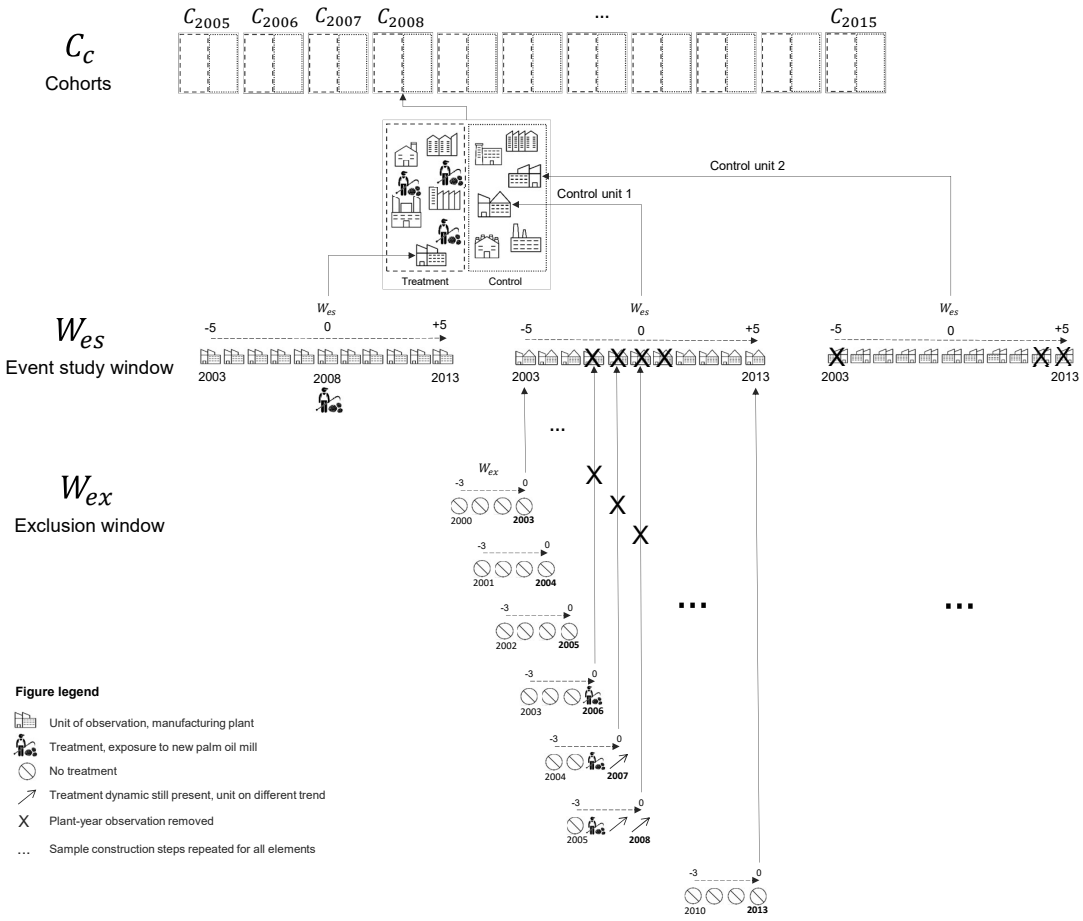
Instead of estimating a standard two-way fixed effects model regressing manufacturing plant outcomes on a running total of palm oil mills in each district or the corresponding total palm plantation area, we therefore use the establishment of a new palm oil mill as a treatment event. Specifically, we identify our treatment effect by comparing manufacturing plants whose treatment status changed in the respective year, which de Chaisemartin and D'Haultfoeuille (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 2.1. Each year within our study period (2005–2015) defines a cohort of palm oil mills (see first row in Figure 2.1). 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.

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

[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.

Figure 2.1: 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 2.5).

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:

$$\ln Y_{\text{idrycs}} = \lambda_i + \mu_{ry} + \nu_{sy} + \delta_0 \text{Treated}_{dc} \times C_c + \Sigma_\tau \delta_1 D^\tau_{cy} \times C_c + \Sigma_\tau \delta_\tau \left(\text{Treated}_{dc} \times D^\tau_{cy}\right) + \epsilon_{\text{idrycs}}, \quad (2.1)$$

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

[6]See Appendix 2.6 for a more extensive discussion and additional applied examples of this design.

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 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):

$$\ln Y_{\text{idrycs}} = \lambda_i + \mu_{ry} + \nu_{ys} + \beta_0 \text{Treated}_{dc} \times C_c + \Sigma_\tau \beta_1 D^\tau_{cy} \times C_c + \beta\left(\text{Treated}_{dc} \times \text{Post}_{cy}\right) + \kappa\left(\text{Treated}_{dc} \times \text{Zero}_{cy}\right) + \epsilon_{\text{idrycs}} \quad (2.2)$$

[7]We also run specifications with fixed effects at the island-industry-year level (see Table 2.3).

[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.

[9]We also show results with a more sparse set of fixed effects in Figure 2.5.

[10]See Appendix 2.6 for a detailed discussion of treatment effect heterogeneity with regards to regression weights in dose-response two-way fixed effects specifications.

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.

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 (2019a) 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 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 2.9 for a breakdown of the number of mills

[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 2.6, 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).

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 one day (Corley and Tinker 2016). 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 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.

[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.

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: One can suspect them of experiencing 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, i.e., excluding the treated districts from the sample.

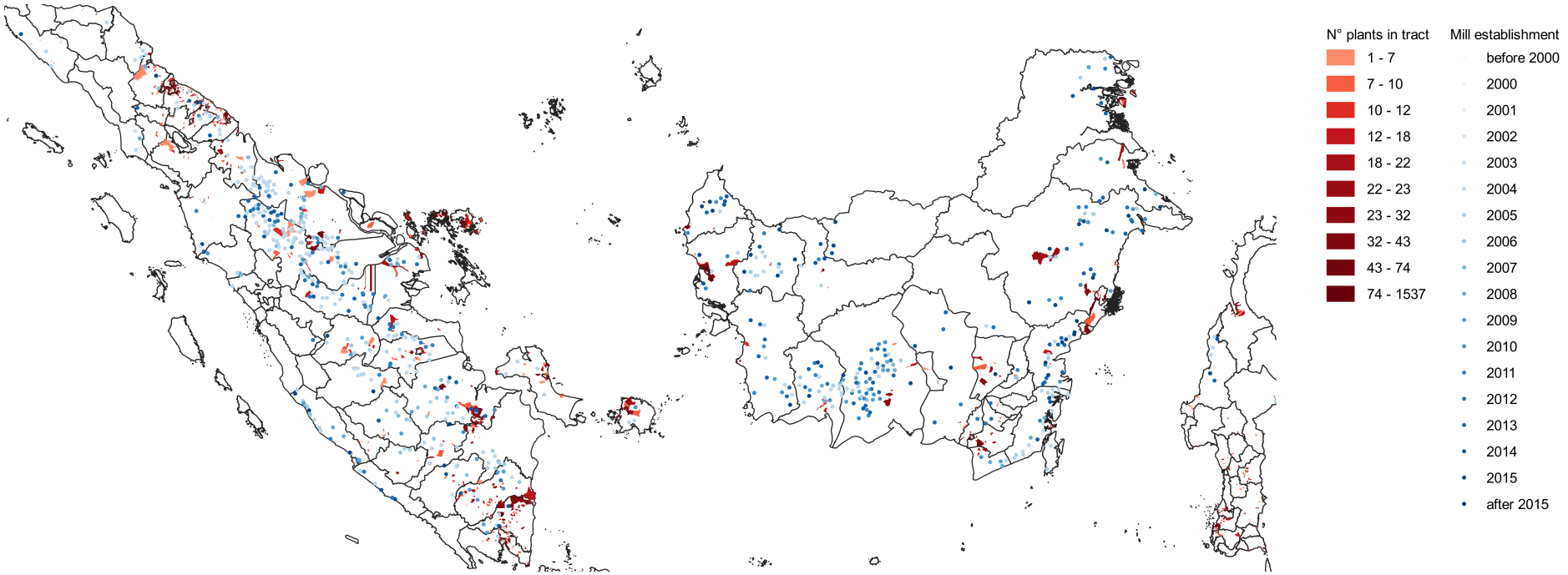
2.3 Data

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,

[16]See Appendix 2.6 for additional detail on the construction of the individual variables.

Figure 2.2: 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.

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.

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 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.

Table 2.1: 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.

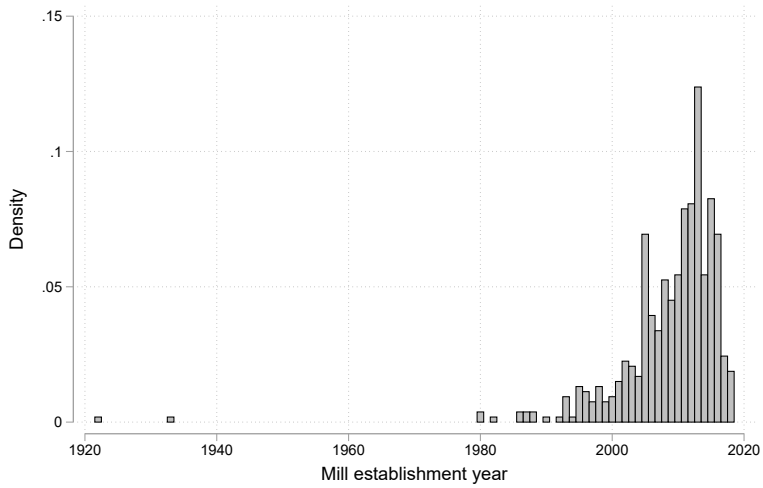
[17]The base sample of our panel is a merge of existing mill location data from researchers and NGOs, the "universal mill list" (World Resources Institute et al. 2019). Data comes from World Resources Institute (WRI), Rainforest Alliance, Proforest, Daemeter, Trase, Earthworm Foundation, Auriga, CIFOR, 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`

[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 2.6 and Figures 2.5, 2.6, and 2.7 for the corresponding coefficients.

Figure 2.2 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 2.1, 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 2.3. 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 2.9, we show a breakdown of the number of mills of corporate groups by islands.

Figure 2.3: 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.

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 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 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),

[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 sample size they generate creating an incentive for them 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.

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 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 2.7) to different methods from the production function estimation literature (Ackerberg et al. 2015; Levinsohn and Petrin 2003; Wooldridge 2009).[24]

Figure 2.2 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 2.2 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.

[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`

[24]See Appendix 2.6 for more detail on the estimation of our production functions.

Table 2.2: 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

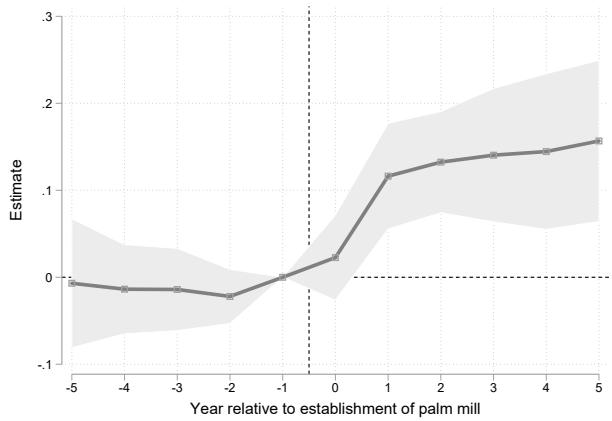
Notes. This table reports summary statistics for outcomes and auxiliary variables (for production function estimations) at the manufacturing plant level. Data is from the Indonesian manufacturing census, which includes a total of 18071 plants in 1994, 19887 in 2005, and 22416 in 2015 (excluding palm 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).

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 (The World 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.

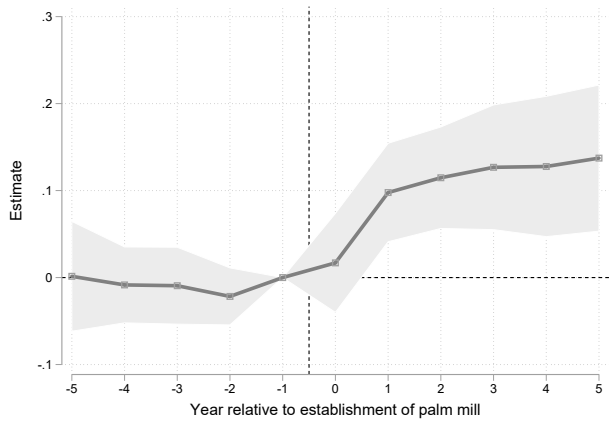
2.4 Results: Local agglomeration effects of palm oil mill shocks

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 mill establishment over the preceding three years. Figure 2.4 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.

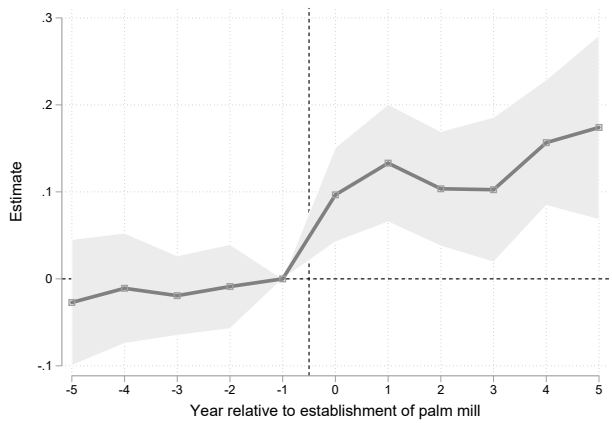
Figure 2.4: Dynamic effects of palm oil mill establishment on sales, labor productivity and total factor productivity



(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 2.1 and are also shown in Table 2.3. 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.

Table 2.3 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 2.1 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.

Table 2.4 shows estimates from Equation 2.2. 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 2.12).

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 2.5). 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 (see Figure 2.5). 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

[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 2.13 shows estimates for the restricted sample of linked plants, which lack statistical power.

Table 2.3: 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 2.1. They are also shown in Figure 2.4. 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 2.2 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 2.4: 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 2.2.
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 2.2 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).

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 2.5 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.

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 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

Table 2.5: 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 2.2 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).

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 2.4).

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).

Road infrastructure and district budgets Government revenues and the quality of local transportation infrastructure improved after the establishment of new palm oil mills. Table 2.6 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.

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

[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.

Table 2.6: 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 2.2 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.

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 2.5) do not change the magnitude of our main effects (reported in Table 2.4). 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 2.4, 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 2.7 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.

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 2.5 and Figures 2.6, and 2.7 in the Appendix (Section 2.6)) that compare point estimates from standard dose-response fixed

Table 2.7: 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 2.2.
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).

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.

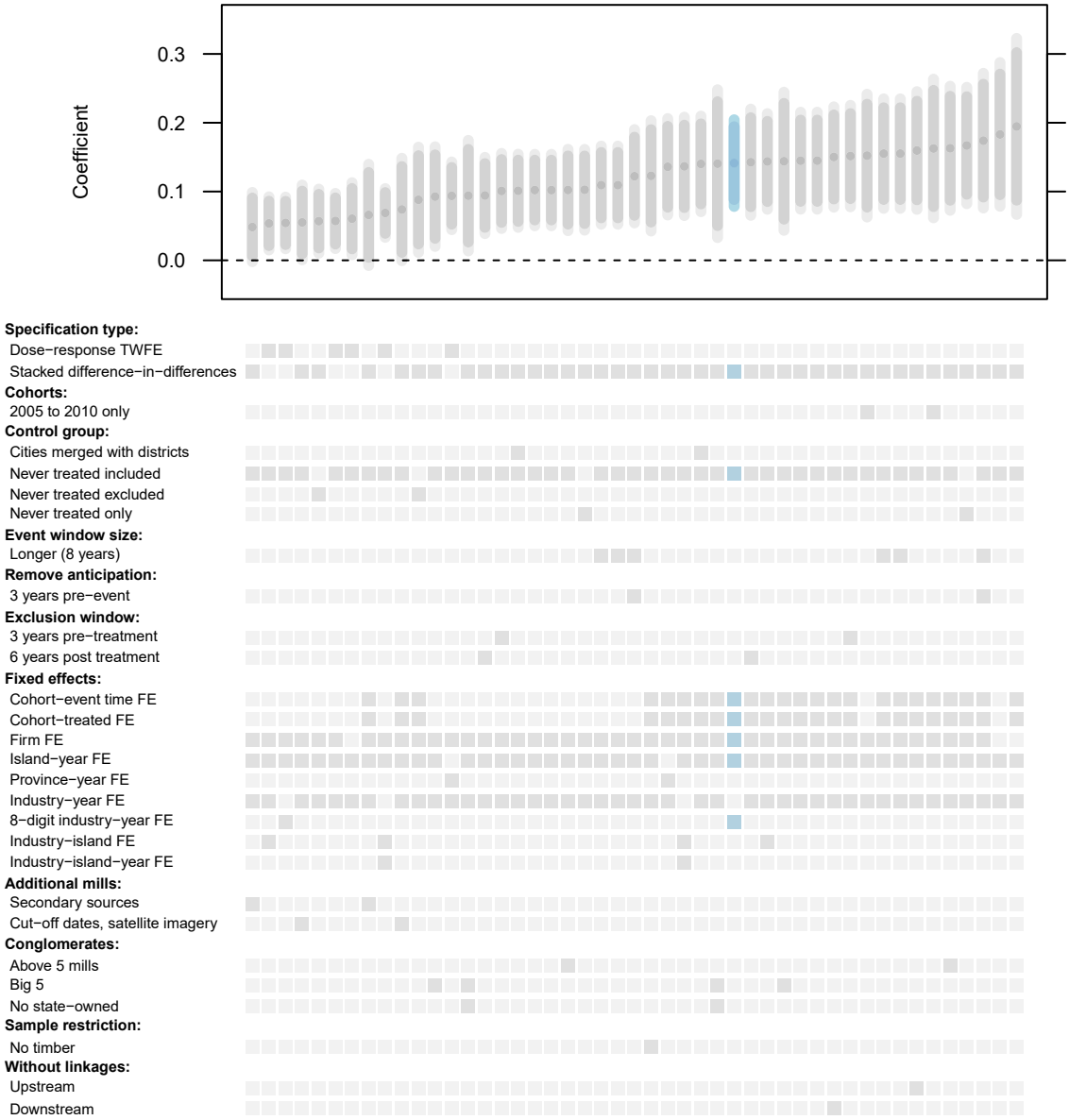
Estimates of a standard two-way fixed effects panel regression with the count of mills in each district as the treatment variable (see Figure 2.5) show effect sizes of similar magnitudes. Goodman-Bacon (2018) shows that any difference could come from two sources: either common trends violations or the weighted aggregation of heterogeneous effects between cohorts and between groups of different treatment.[28]

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

[28]See Appendix 2.6 for a detailed discussion of the difference between our stacked design and standard dose-response two-way fixed effects specifications.
[29]See Appendix 2.6.

Figure 2.5: 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 2.2) 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.

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 2.9, 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 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 2.8). The coefficient estimates are small in magnitude and statistically insignificant (see Table 2.2). The lack of evidence for spillovers in this setup supports the tenability of the SUTVA.

Table 2.8: 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 2.2.
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 2.2 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).

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.

2.5 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.

2.6 Appendix

Additional robustness checks

Figures 2.6 and 2.7 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 2.5). 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.

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.

Adding data sources on mill establishment dates Our main estimation sample is based on establishment dates collected from administrative records (see 2.3 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.

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.

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.

Empirical methods background

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).

Regression weights and heterogeneous treatment effects Besides mitigating violations of the parallel trends assumption, 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'Haultfoeuille 2017).

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 oil 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 2.6).

Data cleaning

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 2.6) 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').

Treatment: Palm oil mill panel

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.

Constructing the stacked data set We manually create data sets for each cohort of palm oil mill establishments, according to the rules described in 2.2. 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 2.2, 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.

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

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

2.10 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 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 2.6). Also, the statistics agency BPS has changed its district coding system several times during our

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

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 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.

Total factor productivity TFP estimates are shown in Table 2.11. 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.

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

Background information

Palm oil industry

Elasticity of palm oil demand Palm oil is a substitute for other vegetable oils and even for petroleum diesel in some 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.

Institutional background

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

[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.

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).

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 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

[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).

[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

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.

[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.

Appendix tables

Table 2.9: 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 oil 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 2.10: 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 2.11: 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 2.12: 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 2.2.
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 2.2 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.

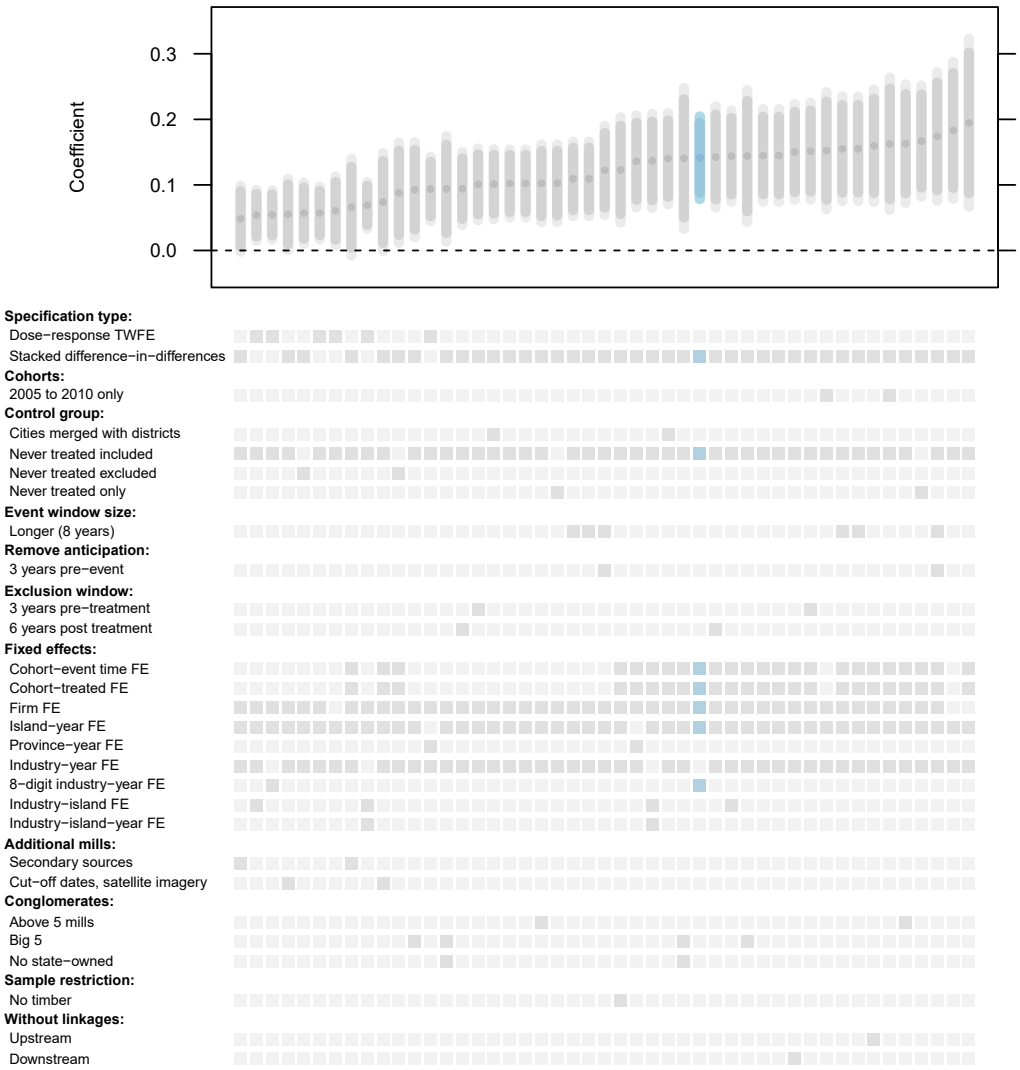
Table 2.13: Effects of palm oil mill establishment on downstream and upstream local manufacturing plant performance

	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 2.2.
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 2.2 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).

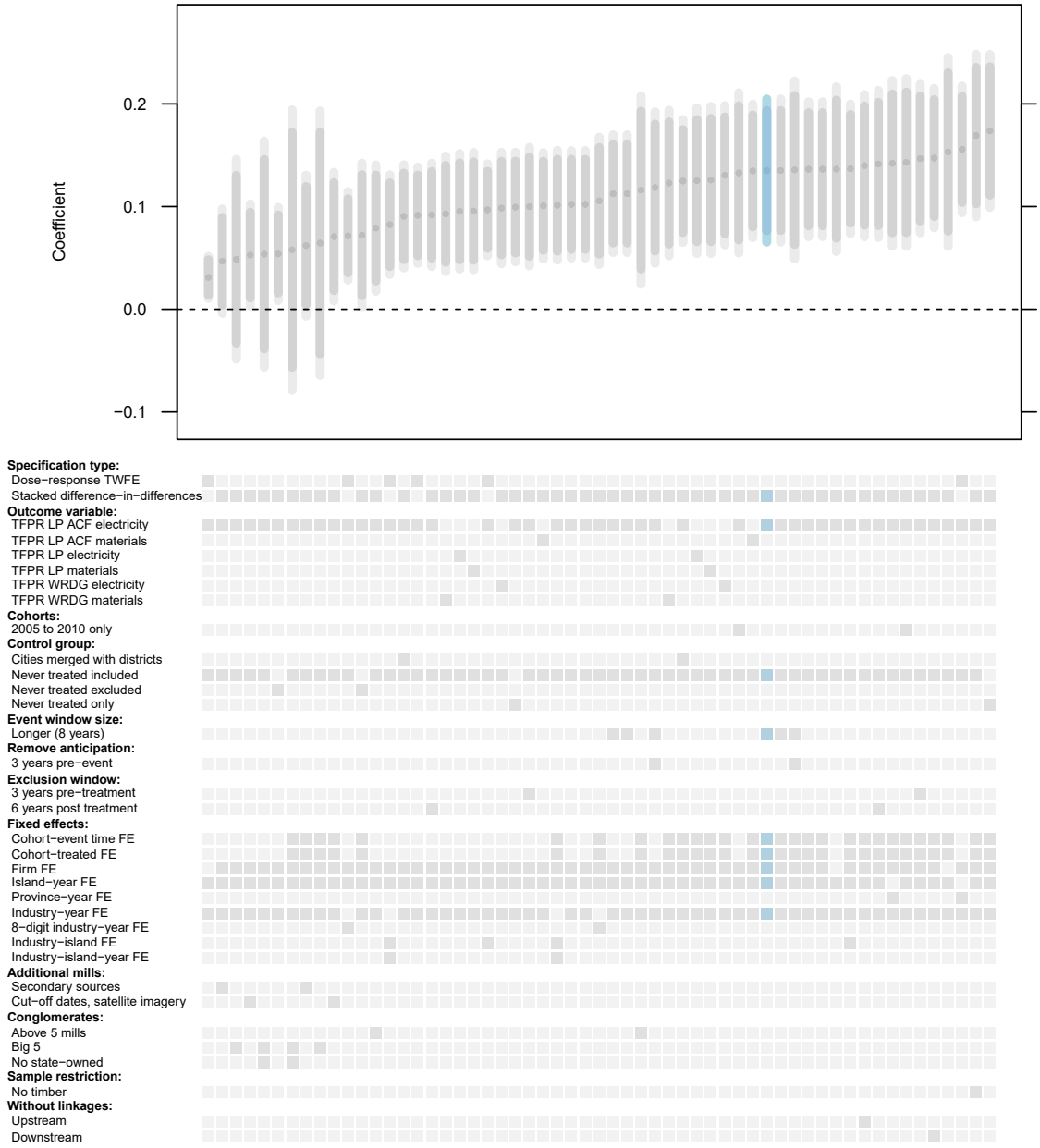
Appendix figures

Figure 2.6: 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 2.2) 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 2.7: 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 2.2) 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.

2.7 References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer (2015). "Identification Properties of Recent Production Function Estimators". *Econometrica* 83.6, pp. 2411–2451. DOI: 10.3982/ECTA13408.

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.

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. DOI: 10.1093/restud/rdx042.

Athey, Susan and Guido Imbens (2018). *Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption*. Working Paper 24963. National Bureau of Economic Research. DOI: 10.3386/w24963.

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

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).

Baudoin, Alice, Pierre-Marie Bosc, Cécile Bessou, and Patrice Levang (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. DOI: 10.3386/w24625.

Bellemare, Marc F. and Jeffrey R. Bloem (2018). "Does Contract Farming Improve Welfare? A Review". *World Development* 112, pp. 259–271. DOI: 10.1016/j.worlddev.2018.08.018.

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. DOI: 10.1016/j.jdeveco.2004.06.004.

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, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber (2012). "The Political Economy of Deforestation in the Tropics". *The Quarterly Journal of Economics* 127.4, pp. 1707–1754. DOI: 10.1093/qje/qjs034.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli (2016). "Agricultural Productivity and Structural Transformation: Evidence from Brazil". *American Economic Review* 106.6, pp. 1320–1365. DOI: 10.1257/aer.20131061.

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

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs". *The Quarterly Journal of Economics* 134.3, pp. 1405–1454. DOI: 10.1093/qje/qjz014.

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. DOI: 10.1002/9781118953297.ch15.

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. DOI: 10.1080/00074918.2016.1184745.

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

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. DOI: 10.1086/705547.

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'Haultfoeuille (2017). "Fuzzy Differences-in-Differences".

De Chaisemartin, Clément and Xavier D'Haultfoeuille (2018). "Fuzzy Differences-in-Differences". *The Review of Economic Studies* 85.2, pp. 999–1028. DOI: 10.1093/restud/rdx049.

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. DOI: 10.1093/restud/rdz017.

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. DOI: 10.1257/pol.20180076.

Edwards, Ryan B. (2019a). "Export Agriculture and Rural Poverty: Evidence from Indonesian Palm Oil". Mimeo.

– (2019b). "Spillovers from Agricultural Processing". Mimeo.

Euler, Michael, Vijesh Krishna, Stefan Schwarze, Hermanto Siregar, and Matin Qaim (2017). "Oil Palm Adoption, Household Welfare, and Nutrition Among Smallholder Farmers in Indonesia". *World Development* 93, pp. 219–235. DOI: 10.1016/j.worlddev.2016.12.019.

Faber, Benjamin and Cecile Gaubert (2019). "Tourism and Economic Development: Evidence from Mexico's Coastline". *American Economic Review* 109.6, pp. 2245–2293. DOI: 10.1257/aer.20161434.

Fadlon, Itzik and Torben Heien Nielsen (2015). *Family Labor Supply Responses to Severe Health Shocks*. Working Paper 21352. National Bureau of Economic Research. DOI: 10.3386/w21352.

– (2019). "Family Health Behaviors". *American Economic Review* 109.9, pp. 3162–3191. DOI: 10.1257/aer.20171993.

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.

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, Rosyani Asnawi, and Matin Qaim (2017). "Oil Palm Boom, Contract Farming, and Rural Economic Development: Village-Level Evidence from Indonesia". *World Development* 95, pp. 127–140. DOI: 10.1016/j.worlddev.2017.02.013.

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. DOI: 10.1016/j.landusepol.2015.03.001.

Goodman-Bacon, Andrew (2018). *Difference-in-Differences with Variation in Treatment Timing*. Working Paper 25018. National Bureau of Economic Research. DOI: 10.3386/w25018.

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. DOI: 10.1093/rfs/hhr011.

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

Heilmayr, Robert, Kimberly M. Carlson, and Jason Jon Benedict (2020). "Deforestation Spillovers from Oil Palm Sustainability Certification". *Environmental Research Letters*. DOI: 10.1088/1748-9326/ab7f0c.

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. DOI: 10.1086/674633. http://users.econ.umn.edu/~holmes/data/plantsize/index.html.

Jelsma, Idsert, G. C. Schoneveld, Annelies Zoomers, and A. C. M. van Westen (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. DOI: 10.1016/j.landusepol.2017.08.012.

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

Kelley, Lisa C., Nancy Lee Peluso, Kimberly M. Carlson, and Suraya Afiff (2020). "Circular Labor Migration and Land-Livelihood Dynamics in Southeast Asia's Concession Landscapes". *Journal of Rural Studies* 73, pp. 21–33. DOI: 10.1016/j.jrurstud.2019.11.019.

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. DOI: 10.1111/1467-937X.00246.

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. DOI: 10.1016/j.geoforum.2017.10.012.

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. DOI: 10.1080/03066150.2010.512460.

Moscona, Jacob (2019). "Agricultural Development and Structural Change Within and Across Countries". Mimeo.

Olley, G. Steven and Ariel Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64.6, pp. 1263–1297. DOI: 10.2307/2171831.

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. DOI: 10.1146/annurev-resource-100815-095459.

Paoli, Gary D., Piers Gillespie, P.L. Wells, L. Hovani, A.E. Sileuw, N. Franklin, and Jim Schweithelm (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. DOI: 10.5129/001041516817037754.

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. DOI: 10.1080/00074918.2016.1228829.

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, Kibrom T. Sibhatu, Hermanto Siregar, and Ingo Grass (2020). "Environmental, Economic, and Social Consequences of the Oil Palm Boom". *Annual Review of Resource Economics* 12.1. DOI: 10.1146/annurev-resource-110119-024922.

Rauch, James E. (1999). "Networks versus Markets in International Trade". *Journal of International Economics* 48.1, pp. 7–35. DOI: 10.1016/S0022-1996(98)00009-9. 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).

The World Bank (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).

Wooldridge, Jeffrey M. (2009). "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables". *Economics Letters* 104.3, pp. 112–114. DOI: 10.1016/j.econlet.2009.04.026.

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

World Resources Institute, Rainforest Alliance, Proforest, Daemeter, Trase, Earthworm, Auriga, CIFOR, Transitions, Jason Jon Benedict, Robert Heilmayr, and Kimberly M. Carlson (2019). *Universal Mill List*. URL: https://data.globalforestwatch.org/datasets/5c026d553ff049a585b90c3b1d53d4f5_34 (visited on 12/14/2020).

Chapter 3

Provisional COVID-19 infrastructure induces large, rapid increases in cycling[†,‡]

Abstract

The bicycle is a low-cost means of transport linked to low risk of transmission of infectious disease. During the COVID-19 crisis, governments have therefore incentivized cycling by provisionally redistributing street space. We evaluate the impact of this new bicycle infrastructure on cycling traffic using a generalized difference-in-differences design. We scrape daily bicycle counts from 736 bicycle counters in 106 European cities. We combine this with data on announced and completed pop-up bike lane road work projects. Within four months, an average of 11.5 km of provisional pop-up bike lanes have been built per city and the policy has increased cycling between 11% and 48% on average. We calculate that the new infrastructure will generate between $1 and $7 billion in health benefits per year if cycling habits are sticky.

3.1 Introduction

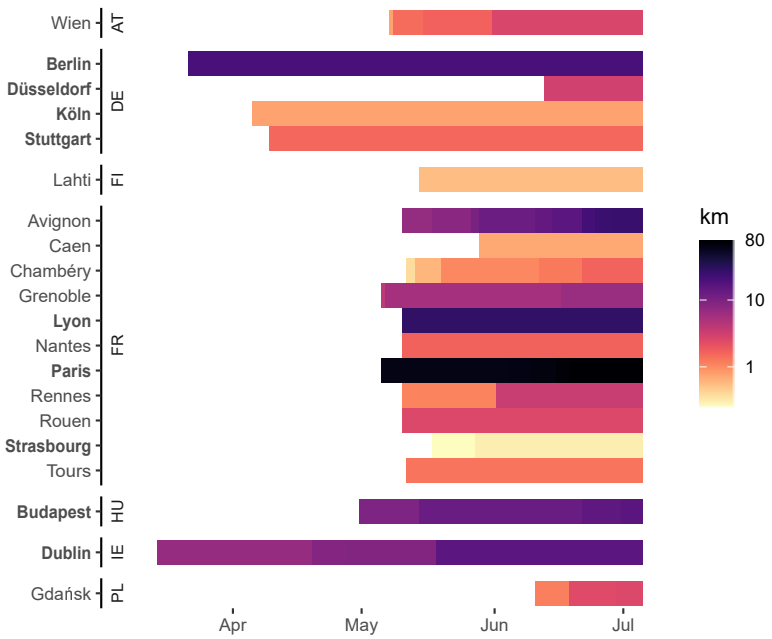
The COVID-19 crisis has led to important changes in transport behavior in 2020 (Apple 2020). Early evidence points to shifts from public transport to car use as users have reacted to the pandemic (Chang et al. 2020). Governments have incentivized cycling as a low-cost, sustainable, equitable, and space-saving mode of transport that reduces the risk of COVID-19 transmission. A key measure has been the redistribution of street space in cities to create provisional bike infrastructure typically marked and protected by materials readily available from road construction companies. As of 8 July 2020, 2,000 km of these infrastructure changes had been announced in European cities (ECF 2020).

Transport mode choices are influenced by a variety of behavioral effects that make people stick to their habits, such as status quo bias, default effects, and time-inconsistent preferences (Mattauch et al. 2016). This complicates the task of policy-makers to encourage people to cycle, particularly in the short-run. However, major disruptions to public transport, such as strikes, cause people to reconsider their habits (Larcom et al. 2017) and the provision of dedicated infrastructure has been identified as an important means to increase cycling (Pucher et al. 2010). Thus, the fast provision of new bike infrastructure during the COVID-19 pandemic is a suitable policy experiment to investigate the responsiveness of cycling under conducive conditions.

[†]Sebastian Kraus and Nicolas Koch (2021). "Provisional COVID-19 Infrastructure Induces Large, Rapid Increases in Cycling". *Proceedings of the National Academy of Sciences* 118.15. DOI: 10.1073/pnas.2024399118

[‡]We thank Ben Thies and Lennard Naumann for their excellent research assistance.

Figure 3.1: Treated cities and their treatment intensities in terms of implemented kilometers of public bike lanes in service (cumulative) on a given day between March and July 2020



Notes. Cities used in the estimation sample for Figure 3.3 are marked in bold. Control cities are plotted in Figure S2 of the SI. London, Milan, Rome, and Lisbon are missing from the sample due to a lack of daily bicycle counter data. Data is from the European Cyclists' Federation (ECF 2020).

Here, we estimate the causal effect of the post-COVID-19-lockdown roll-out of provisional ("pop-up") bike lanes in European cities. We compile new data on daily bike counts in 106 cities. We connect to the open data application programming interfaces (APIs) of these cities to download bike counts from a total of 736 counters. We combine this data with information on day-to-day kilometer changes in pop-up cycling infrastructure (Figure 3.1).

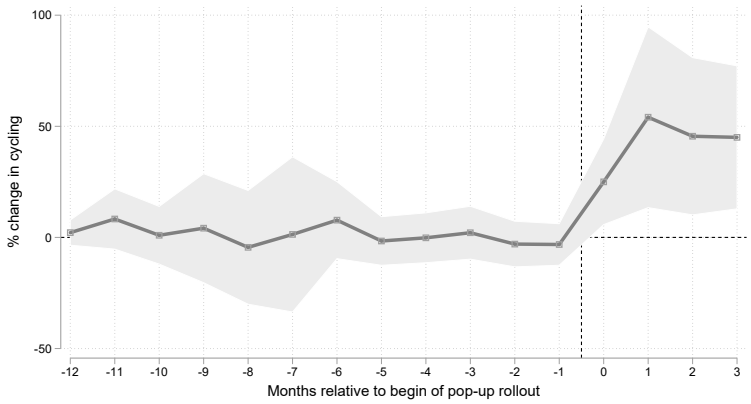
The spatial placement of pop-up bike lanes has mainly been driven by the availability of street space that could be redistributed without restricting car traffic to one direction and the existence of "shovel-ready" construction plans. The exact timing of pop-up bike lane construction is driven by administrative idiosyncrasies and the availability and schedules of construction firms. Therefore, the timing of the pop-up bike lane roll-out has been as good as random. This quasi-experimental setting allows us to address the important concerns that bike lanes could be built as a reaction to increased cycling traffic (reverse causality) or that both the implementation of bike lanes and bicycle counts could be driven by a third factor, such as local "green" preferences, that cannot be measured (omitted variable bias).

3.2 Results

We use panel regressions to compare bike traffic in treated cities before and after they get treated with control cities. We find that pop-up bike lanes have led to substantial increases in cycling. This effect is robustly visible in comparisons over both a longer and a shorter time span. First, in Figure 3.2 we show the effect comparing treated and control cities over several months before and after treatment. Second, in Figure 3.3 we provide estimates from a range of more conservative specifications identifying the effect based on daily variation within a narrow time window in the same city.

The outcome in all our regressions is modeled as the natural logarithm of the cycling count. We use daily variation in this variable either at the counter or at the city-level. Our coefficients can be interpreted as the average change in cycling caused by the pop-up bike lane program.

Figure 3.2: Treatment effect (difference between treated and control cities) in months before and after the beginning of the pop-up bike lane policy



Notes. Observations are binned into months. Treatment for this plot is hard-coded to March 2020 and the baseline category and the beginning of the sample are set to February 2019. Estimates are from Poisson regressions that include city and country-day fixed effects (Equation 3.3, SI Appendix, Section 3.5). The shaded area shows the 95% confidence interval. Data for the outcome variable is from the European Cyclists' Federation (ECF 2020) and data for the treatment variable is from municipal bike counters (Materials and Methods).

Standard difference-in-differences

Figure 3.2 shows the dynamic treatment effect of the pop-up bike lane program. For the analysis shown here, we define March 2020 as the time of treatment and plot the estimated differences between treated and control cities over time. Since we expect cycling to increase in both treated and control cities as a reaction to COVID-19, we take the difference between the cycling increase in treated and in control cities as our estimate of the average effect of the program. This difference-in-differences approach suggests an increase in cycling of 41.6% induced on average by the policy. A crucial assumption for this research design is that cycling would have evolved on a parallel trend in the treatment and control group in

the absence of treatment. This is called the common trends assumption. Since we model the outcome as the natural logarithm of cycling counts, we make the assumption that cycling would have grown at the same rate in the treatment and in the control group.

Figure 3.2 allows us to verify this assumption. The treatment effect becomes apparent after the treatment sets in. Before, treatment and control group have been on the same trend. There is a slight, albeit statistically insignificant downward trend before treatment, hinting at the possibility of stronger mobility reductions due to COVID-19 in cities that have decided to build pop-up bike lanes. This could for instance be the case because local and national governments are more likely to take wide-ranging action if their country is hit by a more intense outbreak. It could also be due to governments acting upon stronger risk-aversion of their local populations towards cycling in the context of emptier roads and increased speeding during the lockdown. We mitigate some of these potential selection into treatment effects by controlling for COVID-19 related dynamics with fixed effects at the country-day level. This removes the effect of daily national level policy changes, such as lockdowns.

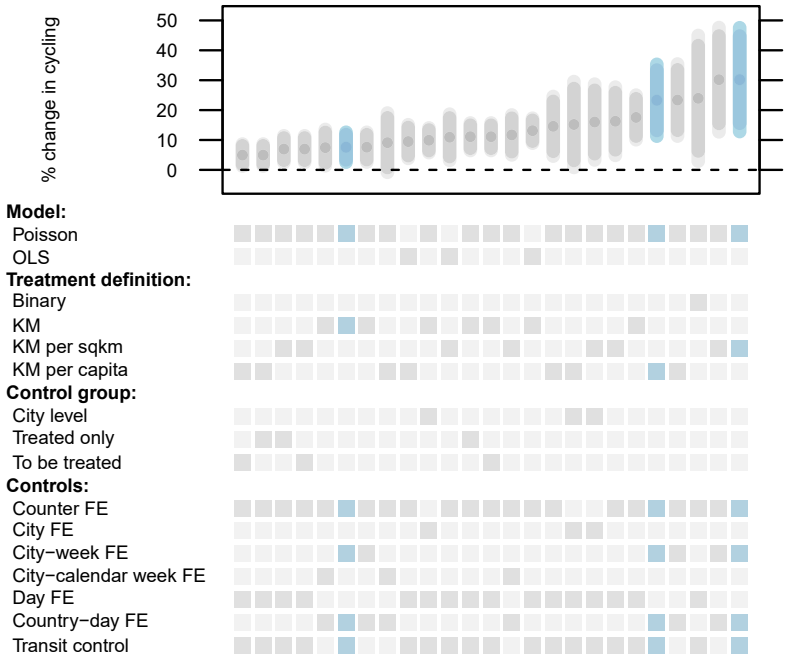
A remaining concern is that bike lanes could have been built as a reaction to locally increased cycling traffic (reverse causality) or that both the implementation of bike lanes and bicycle counts could be driven by an unaccounted third factor (omitted variable bias). We address these potential biases with regressions focusing on changes over a shorter time span as discussed in the next section.

Generalized difference-in-differences

In our second set of specifications (Figure 3.3) we investigate more focused comparisons using both variation in the timing of treatment between cities and in the treatment dose, i.e. the number of kilometers of pop-up bike lane in service on a given day. With these specifications we robustify the more simple difference-in-differences design by using additional fixed effects and by including control variables for the weather, for changes in overall mobility and public transport, and for the number of active bike counters in a city. Crucially, we look at the effects of pop-up bike lanes in a shorter time span to investigate potential reverse causality between cycling and the implementation of pop-up bike lanes. Although pop-up bike lanes tend to be based on pre-existing plans by city planners or civil society organizations and could therefore be implemented comparatively fast, the erection of a bike lane needs at least a few days notice and the *exact* timing of these road works depend on the availability and the schedule of construction firms. This has been confirmed in our conversations with local policy-makers in Berlin and Paris. Our preferred specifications (Equation 3.1) are therefore based on comparisons of cycling counts on the days before and after a change in the treatment intensity (marked in blue). These comparisons are created by the inclusion of city-week fixed effect. This fixed effect ensures that our estimates are based on variation within the same city within the same

week. If the exact (i.e., day-level) timing of the roll-out of pop-up bike lanes has been as good as random, estimates from these specification are not driven by reverse causality.

Figure 3.3: Estimates of the average effect of pop-up bike lanes on cycling. Dose-response regressions (in km, km per capita, or km per km^2 in service on a given day) are multiplied by the average treatment dose



Notes. The unit of observation is the bike counter except for regressions at the city level. Preferred specifications are marked in blue (Equation 1) and are reported in more detail in Table 3.6 (SI Appendix, Section 3.5). Darker colors in the bottom panel indicate the type of specification. The 90% and 95% confidence intervals are shown in darker and lighter color. Three estimates are from OLS specifications and therefore use the natural logarithm of the bicycle count as the outcome. All other specifications are Poisson regressions using the level of the count. Data for the outcome is from the European Cyclists' Federation (ECF 2020) and data for the treatment is from municipal bike counters (Materials and Methods). All regressions include controls for the number of active counters in a city on a given day and for the weather (temperature, sunshine, wind, precipitation) (Hersbach et al. 2020). All regressions, except those that rely on observations before 2020, include a control for overall mobility (Facebook 2020). The transit control is from Apple routing requests (2020 only) (Apple 2020). Code for the figure is from (Ortiz-Bobea 2020).

Our unit of observation in most regressions is the cycling counter. This allows us to control for within-city differences despite doing a cross-city study. We do this by including a counter fixed effect that flexibly controls for any local confounders that are time-invariant within the time frame of the variation used in the analysis. We thereby control for the density of public transport stops, population density, topography, but also additional, unobservable dimensions, such as social capital and local preferences for green lifestyles, at a high spatial resolution within the city. With the counter fixed effect we also rule out that our result is driven by new counters that get placed next to pop-up bike lanes. We assign treatment to each counter based on its city, since we measure daily changes in the pop-up bike lane network at that level. We investigate the effect of this source of measurement error by defining the treatment dose either as a binary variable or in terms of km, km per capita, and km per km^2 city area. We find that measuring the dose-response in terms of kilometers attenuates the effect (7.6%). This indicates that the effect is not exclusively driven by the announcement effect of new infrastructure in a city, but by the de

facto availability of new infrastructure in the neighborhood surrounding a counter, which is better approximated by a measure in per capita or per area terms (estimates of 23.3% and 30.2%, respectively). Remaining measurement error due to some counters being closer or farther from new infrastructure than the rest of the sample is unlikely to be systematic conditional on fixed effects and control variables (detailed discussion of measurement error in SI Appendix, Section 3.5). We also run specifications for which we take the mean of all counters in a city (marked as *City level* in Figure 3.3) to show that the effect is not driven by our use of the counter as the unit of observation.

We use a variable capturing transit routing searches on Apple maps (Apple 2020) to control for omitted variable bias that could be present if changes in public transport affect both pop-up bike lane construction and cycling. In our preferred specification this could still be the case, if *daily* changes in the provision or in the use of public transport in a city led to new pop-up infrastructure *within the same week*. Public officials may for instance have tried to schedule the erection of pop-up bike lanes for the same day as planned public transport disruptions. The transit control removes this potential remaining bias. Since the Apple data is only available for a subset of larger cities in our sample (marked in bold in Figure 3.1), we run our main regressions (Figure 3.3) on this smaller sample. Table S4 (SI) shows robustness to lifting this sample restriction and to excluding Paris, which has had the strongest treatment, from the analysis.

We control for sub-national changes in policies and behaviors related to COVID-19 with a variable that captures overall human mobility based on Facebook user movements. We control for the number of counters active in a city on a given day to account for unusual traffic situations, for instance when a counter gets shut down because of road works. We also include control variables for daily total precipitation and mean wind, temperature, and sunshine to address the concern that both the scheduling of pop-up bike lane construction work and daily variation in cycling could have been driven by weather conditions.

We check the sensitivity of our results to changing the time-span of our identifying variation and to reshaping our treatment and control group definitions (additional specifications in Figure 3.3). The effect is robust to including days from the same calendar week in previous years in these comparisons rather than days from 2020 only. We also provide estimates for the effects of the policy based on comparisons between (i) treated and untreated cities, (ii) treated cities only using their variation in treatment timing, (iii) cities that are already treated and those that have only announced pop-up bike lanes, (iv) treated cities only using their variation in treatment dose and treatment timing.

Heterogeneity analysis

We investigate how the treatment effect of pop-up bike lanes varies depending on relevant features of the cities in our sample (Table 3.1). These heterogeneous effects should not

be interpreted causally, since we cannot control for additional omitted variable bias or reverse causality created by the inclusion of these variables in our model. We find that the effect of pop-up bike lanes is stronger in cities with a higher population density (1) and a higher modal share of public transport in commutes (2), which are correlates of a built-environment favoring active travel. The treatment effect is lower for cities with faster average speeds of car commutes (6) and for cities with more road death per capita (7). It is also lower for cities with more cars per capita (5). However this estimate is imprecise. These heterogeneities confirm research that found that US cities with better safety, low car ownership, and more density have more cycling (Buehler and Pucher 2012; Buehler et al. 2020).

Our analysis also shows that the length of the bike lane network per capita (3) is correlated with a lower treatment effect. We interpret this as an indication that the pop-up bike lane effect is a phenomenon of catch-up growth in cities with a high cycling potential, that was previously impeded by missing infrastructure. The effect of baseline cycling modal shares (4) is, however, statistically unclear.

Further research could also look at the effect of pop-up bike lanes in terms of improvements in bike lane network connectivity and directness as proposed by (Schoner and Levinson 2014) and other more complex measures of a bike lane network, such as the level of protection of a bike lane and the treatment of intersections (Buehler and Dill 2016). In this context it is important to investigate, how underserved communities can be provided with a pop-up bike lane network, that is complete and inclusive, and how additional political, cultural, and economic barriers to cycling for low-income and minority groups can be removed (Agyeman 2020). Bike sharing can support changes in modal choice (Hamilton and Wichman 2018), but important barriers to adoption remain for underrepresented groups (McNeil et al. 2018). We therefore think it would be valuable to investigate interactions between the pop-up bike lane policy and time series data on bike sharing policies including changes in pricing and the availability of bikes and stations.

Table 3.1: Heterogeneous treatment effects of the pop-up bike lane roll-out

	× baseline (natural log) of						
	(1) Population density	(2) PT modal share	(3) Bike lanes p/c	(4) Cycling modal share	(5) Cars p/c	(6) Car speeds	(7) Road deaths p/c
Pop-up	0.221*	0.258***	-0.194*	0.093	-0.592	-0.509**	-0.351***
	(0.121)	(0.100)	(0.115)	(0.082)	(0.485)	(0.233)	(0.058)
N	59521	27486	24611	27486	34408	26886	34922

Notes. Estimates are based on the interaction term of the treatment variable (in km per city area) and the natural logarithm of the heterogeneity variables (column names). Coefficients are scaled to the average treatment dose in our sample. They can be interpreted as the unit change in cycling if a heterogeneity variable is one unit higher (when assuming treatment with an average pop-up bike lane program). All regressions include counter, city-week, and country-day fixed effects. They also include weather controls (Hersbach et al. 2020), a control for overall mobility (Facebook 2020), and a control for the number of counters active in a city on a given day. Data for the outcome variable is from the European Cyclists' Federation (ECF 2020) and data for the treatment variable is from municipal bike counters (Materials and Methods section). All heterogeneity variables except for *Bike lanes p/c* (Mueller et al. 2018) are from the European Urban Audit (Eurostat 2020a,b). Standard errors clustered at the city level are reported in parentheses. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.3 Discussion

We find robust evidence for substantial short-run increases in cycling in European cities due to new provisional cycling infrastructure. Independent of its potential impacts reducing COVID-19 transmission, the net benefits of the intervention are likely to be large. The direct cost of cycling infrastructure including planning is low. At the higher end, one kilometer of bike lane in Sevilla has previously cost €250000 (Marqués et al. 2015). But, Berlin's approach of iterative planning with provisional infrastructure during the pandemic has for instance reduced costs to €9500 per km as of July 2020 (Bezirksamt Friedrichshain-Kreuzberg 2020). These costs are small compared to the substantial health benefits from the new infrastructure. Previous research has found that every kilometer of cycling generates health benefits of $0.45 (Zapata-Diomedi et al. 2018). As a complementary and more stylized analysis, we combine this estimate from the literature with our econometric estimates of policy-induced cycling increases to provide a projection of health benefits generated by pop-up bike lane programs. We calculate baseline values for total cycling in a city based on data on daily kilometers cycled in German cities in 2018 and extrapolate these numbers to the other European cities in our sample based on city-level data on transport and population (Materials and Methods). This extrapolation is approximate but sufficient to calculate a range of potential health benefits. Based on our regression-based estimate for the 95% confidence interval of the "treatment dose" in terms of km per km^2, we project that the additional cycling induced by the pop-up bike lane treatment during its first months of operation has generated between $0.5 and $1.7 billion in health benefits. Thus, the new infrastructure may generate between $2.2 and $6.9 billion per year in health benefits if the new bike lanes become permanent and make cycling habits stick. We project this range to be between $1.2 and $3.5 in annual health benefits if we use our alternative estimate for the 95% confidence interval of the policy effect based on the "treatment dose"

in terms of km per capita.

The magnitude of our estimate is large compared to previous evaluations of new cycling infrastructure improvements that have found statistically unclear or modest effects, typically because of the limited scale of the interventions (Aldred 2019; Winters et al. 2017; Yang et al. 2010). Our estimate implies a higher responsiveness of cycling to new infrastructure than the associations found in cross-sections of US cities (Buehler and Pucher 2012; Dill and Carr 2003). However, in cities in Europe (Mueller et al. 2018) and the UK (Parkin et al. 2008) additional infrastructure is associated with more cycling than in the US. The case of Sevilla has shown that in a dense city with a high share of narrow, cycling-friendly roads the construction of bike lanes on major roads can create substantial cycling growth: 120 km of new bike lane have led there to a five-fold increase of cycling between 2006 and 2011 (Marqués et al. 2015). Similarly, pop-up bike lanes have often been placed on main roads. Thereby they have removed important bottlenecks for cyclists and generated important improvements for the overall cycling network. Many of the cities in our sample are fundamentally well-suited for cycling. For instance, they are often dense and have a high share of side roads with slow car speeds. Therefore, they can be assumed to have a high potential for catch-up growth, which is one explanation our larger effect estimate. In addition, the pandemic has led to a reshuffling of otherwise rather inelastic mobility choices and thus created the conditions for new infrastructure to induce shifts to active travel. However, this also means that our results cannot be directly generalized to other settings. Given this limitation in terms of external validity, we caution against an over-interpretation of our estimates as providing a benchmark value for increases in cycling that planners should expect from an additional kilometer of bike infrastructure. It remains to be evaluated, if the new bicycle use is sticky and how similar treatments influence behavior outside of the context of a pandemic.

Surveys indicate that separated, protected infrastructure is a key element to incentivize up-take of cycling (Aldred et al. 2017; Dill and McNeil 2016; Manaugh et al. 2017). Cities have experimented with different measures to create new spaces for cycling, ranging from painted to provisionally protected bike lanes and from traffic calming with signs to built "modal filters" that prevent the passage of cars. Our data on popup infrastructure does not allow us to systematically distinguish between these types of interventions and the quality of their implementation.[1] Further research should investigate, which types of infrastructure have more successfully increased cycling by previously underrepresented groups, such as women, older people, and children.

[1]In Table 3.4 (SI Appendix, Section 3.5) we show that the results are robust to specifying treatment in terms of (i) the total length of all types of infrastructure, (ii) the total length of measures clearly marked as bike lanes in the data, (iii) the number of measures, and (iv) a binary indicator for treatment.

3.4 Methods

Bicycle counter data

We connect to municipal Open Data Portals to obtain daily bicycle counts from bike counters in large- and medium-sized cities in 20 European countries. The raw data and code to download counter data is included in our code package (Kraus and Koch 2020). The outcome is modeled as the natural logarithm of cycling counts. This means, that we investigate percentage changes rather than absolute increases in the number of cyclists. Our outcome varies daily at the counter-level (summary statistics and cleaning procedures in SI).

Bike lane data

The data on planned and completed pop-up infrastructure projects has been collected and crowdsourced by the European Cyclists' Federation based on technical reports and media announcements. A visualization of the data can be accessed at: `https://ecf.com/dashboard`. We merge this data with city polygons from the European Urban Audit 2020 (Eurostat 2020c) to generate a cumulative measure for the total number of pop-up bike lanes in service in a city on a given day (summary statistics and cleaning procedures in SI). We generate a range of treatment variables (binary, km built, km per capita, km per km^2 city area) and assign this treatment to counters based on their city polygon.

Control variables

Using fixed effects in our regressions, we remove and therefore control for time-invariant differences between cities and between the locations of the individual counters in our data. Therefore, any additional time-invariant control variables at the city and counter level would be redundant in our analysis. We also use fixed effects interacting different spatial levels with time dimensions, thereby controlling for many time-varying observable and unobservable factors. We use additional controls to rule out any bias that may be introduced by time-varying factors below our fixed effect levels.

We control for daily changes in public transport supply and demand with the *transit* variable from the *Apple COVID-19 Mobility Trends Reports* (Apple 2020). This variable captures daily variations in the number of requests for public transport directions on Apple Maps. We access this data using the *covmobility* package (Healy 2020).

We capture average human mobility throughout the phase of the COVID-19 pandemic

starting in March with a human mobility index based on Facebook data (Facebook 2020). The index is from a data set called *movement range maps* that Facebook shares after aggregating individual user movements for humanitarian and research purposes with a reference to the principles outlined by epidemiologists and public health researchers (Buckee et al. 2020). It measures the number of daily 600 meter grid cells visited by Facebook users compared to a baseline in February. For most of our sample the index is aggregated to the state-level, where we use the data. On average, in our sample period daily mobility has been below the February baseline.

We use weather data from the ERA5 climate model that generates hourly measures of surface temperature, UV radiation, precipitation and wind at a $0.25° \times 0.25°$ resolution (Hersbach et al. 2020). We use the ecmwfr package (Hufkens et al. 2019) to aggregate this to the EU Urban Audit city polygons (Eurostat 2020c) at the daily level.

Heterogeneity variables

We analyze heterogeneous treatment effects along seven city-level variables. Bike lanes per capita measures the length of the bike lane network in a city based on Open Street Map data (Mueller et al. 2018; Salmon and Mueller 2017). Population density is from the European Urban Audit (Eurostat 2020a). Public transport (PT) modal share, cycling modal share, cars per capita, car commute speeds, and road deaths per capita are based on city transport statistics from Eurostat (Eurostat 2020b). We use the natural logarithm of these variables in order to obtain the unit change in cycling for a unit change in the respective heterogeneity variable.

Empirical strategy

We estimate a panel regression model at the counter level with daily counts of cyclists as the outcome variable and the number of kilometers (km, km per capita, or km per km^2 city area) of pop-up bike lanes in service in a city on a given day as the treatment. This regression analysis forms comparisons between treatment and control groups before and after treatment for each cohort of new bike lanes and for different treatment intensities (generalized difference-in-differences). This separates the effect of pop-up bike lanes from overall changes in cycling due to COVID-19. We use a set of indicator variables (fixed effects) that remove remaining variation from our estimation sample that would otherwise bias our estimates. Our study design thus allows for systematic differences in the level of bike traffic between treatment and control group, but relies on a common trends assumption, that bike traffic in treated and control cities would have evolved on a parallel trend in the absence of treatment. We cannot observe treated units in their untreated state after treatment (potential outcome). However, we can investigate pre-treatment trends between treated and control cities and check the sensitivity of our estimates to changes

in the control group definition, i.e., in the way we construct the empirical counterfactual (Figure 3.3).

In our preferred specification we model the relationship between cycling traffic and the pop-up bike lane treatment as:

$$\ln \text{Count}_{id} = \beta \text{Bike Lanes}_{cd} + \mathbf{X}_{cd} + \lambda_i + \sigma_{cw} + \varphi_{nd} + \varepsilon_{id} \tag{3.1}$$

where i indexes a counter, c a city, n a country, d a day, and w a week.

λ_i is a counter fixed effect that controls for time invariant factors at a high spatial resolution. σ_{cw} is a city-week fixed effect that controls for week-specific time-varying factors, thereby restricting identifying variation to days before and after treatment within the same week in the same city. φ_{nd} is a country-day fixed effect that captures any daily changes common to all cities in a country.

We cannot include fixed effects for factors that vary at the city-level over time, such as local mobility or weather, since this is the geographical level at which our treatment is measured. $\mathbf{X}_{cd}$ is a vector of control variables that account for these factors. It includes an index for public transport use from Apple (2020), an index for overall mobility based on Facebook data (Facebook 2020), weather variables (temperature, UV radiation, wind, precipitation) (Hersbach et al. 2020), and the number of counters per city active on a given day.

The coefficient of interest is β. It captures the effect of the pop-up bike lane treatment on bicycle counts. Our treatment variable is defined either as a binary indicator for treatment or as the number of kilometers (km, km per capita, or km per km^2 city area) of pop-up bike lanes in service on a given day.

Figures 3.2 and 3.3, and Table 3.1 present the transformed estimate: $100 \times (\exp \beta - 1)$.

Since our outcome is a count variable, we use Poisson pseudo-maximum likelihood regressions (PPML) to estimate this model (Correia et al. 2020). As a robustness check we also use ordinary least squares (OLS) with the natural logarithm of the bicycle count as the outcome (Figure 3.3). We cluster standard errors at the city-level, where treatment is assigned (Abadie et al. 2017).

Calculating the health benefits

We calculate the health benefits by combining our regression estimates of cycling increases for each kilometer of pop-up bike lane with an estimate of the average health benefits of a kilometer cycled ($0.45 converted from 0.62 Australian Dollars), which is lower than typical values from the gray literature (Zapata-Diomedi et al. 2018). Our dose-response

regressions give us the percentage increase in cycling per kilometer of bike lane divided by the city size or city population. For each city in our sample we multiply this effect with the size of its pop-up bike lane program. We then convert this result into additional kilometers cycled in a city based on baseline values of kilometers cycled per person from a detailed transport behavior survey in 135 German cities (Hubrich et al. 2019). We impute values of kilometers cycled for other European cities based on ordinary least squares regressions using information on baseline values of a city's modal split (trips) of commutes, its population density, the length of its initial bike lane network, the modal share of public transport, the number of cars per capita, the average speed of car commuting, and road deaths per capita (more detail in section *Heterogeneity variables*).

Data and code availability

Data and code are available at: `https://zenodo.org/record/4015974` (Kraus and Koch 2020).

3.5 SI Appendix

Data cleaning

Bicycle count data We assemble a new data set of daily bicycle counts from municipal bicycle counters. We connect to national and municipal open data portals for bike counter data sets. We connect directly to the API of those cities that use the Eco-Counter standard (see the data and code repository at `https://zenodo.org/record/4015974`). We also obtain longer time series of bike counts going back to 2012 directly from the mayor's staff for road planning and data in Paris.

Our raw data set contains roughly a million daily counts starting in 2007. We drop the lower and upper percentiles from this raw sample since counters can record very low values, when they are not functioning properly or very high values, when there is a cycling event that drives up counts. We drop the counter 100041252 from Bergen that varies between very low values and some of the highest daily counts in the sample. Our results are robust to keeping these extreme values in the sample. The bulk of the bike counts are from most recent years and we focus most of the comparisons made in our regressions on the years 2019 and 2020. For certain treated cities, such as Paris and Berlin, the raw data already indicates an increase in the annual peak in June 2020 compared to June 2019. However, many of the control cities show a similar pattern (see Figure 3.6). In our regression analyses we find a robust effect of new infrastructures, both when taking the difference in these differences between treatment and control cities, but also when

focusing on variation in treatment timing exclusively (i.e. when discarding control city information) (see Figure 3.3).

Except for robustness checks at the city level, the unit of observation in our regression analyses is the bike counter and counts vary daily. An average counter detects 1457 cyclists per day. The mean number of counters active in the same city on a given day for the counters in the sample is 22.9 (mean over all counter-day observations). The mean size of cities in our sample is 33,000 ha (see Figure 3.2).

Pop-up infrastructure data We use project-level data on provisional infrastructure in European cities as a reaction to the COVID-19 pandemic collected by the European Cyclists' Federation (ECF 2020). In the data we typically see the street, where the project is implemented, its size measured in kilometers, the date of announcement, and the date of implementation. The data also contains the type of project. 80% are categorized as bike lanes and 16% as traffic calming. Our data includes all projects recorded as of July 8, 2020. Our sample does not include infrastructure built after that date and excludes some bigger cities, for which adequate open bike counter data is missing.

We aggregate this data at the city-day level to construct a variable of daily implemented kilometers of pop-up bike lane. We use the city definition and corresponding polygons from the European Urban Audit 2020 (Eurostat 2020c). Typically areas defined by the European Urban Audit include suburbs. For instance, the Paris polygon includes many areas beyond the ring highway that surrounds the municipality of Paris ("Ville de Paris"). This allows us to capture commuting enabled by new bike lanes from the suburbs into the city center, which constitute an important share of all projects (see for instance `https://carto.parlons-velo.fr/#10.13/48.8312/2.5506` for a detailed map of projects in France).

Our estimation sample contains 22 treated cities and 84 control cities, both of which some are dropped from our Poisson regressions depending on the specification because of a lack of variation after removing fixed effects or because we do not have observations for our control variables (the transit (Apple 2020) and overall mobility (Facebook 2020) controls). Dublin and Berlin have been the earliest adopters of pop-up bike lanes in the sample and Paris has been the city with the largest program (see Figure 3.1). We use variation in both timing and the extent of the implemented infrastructure to estimate our effects. We have a large sample from both France and Germany. This allows us to estimate our effect based on within-country variation removing time-varying factors related to the pandemic that could create bias in our estimates. While important cities such as London, Milan, Lisbon and Rome had either announced or already implemented a pop-up bike lane program at the time of the analysis, they are missing from the sample due to insufficient spatial or temporal coverage of the bike count data. The average length by city of all bike infrastructures in our sample combined is 11.5 km, the length of bike lanes is 8.2 and the number of measures implemented 19.8 (see Figure 3.3).

We check the sensitivity of our results to different specifications of the treatment, for instance as an indicator variable that is 1, if there is any cycling related infrastructure change in a city and 0 otherwise (see Figure 3.4).

Measurement error The unit of observation in our preferred specifications (see Equation 3.1 in the Materials and Methods section) is the counter. Our estimates give the average effect over all counters in all cities in the sample. We assign the treatment to counters at the city-level, but our research design ensures that conditional on fixed effects and control variables treatment is as good as random. Therefore, measurement error in our natural experiment can be analyzed similarly to the stylized case of a cluster randomized controlled trial (RCT), for instance with a treatment that is randomized and assigned at the village or class-room level, but outcomes are measured for individuals.

We use different treatment definitions to investigate how different ways of conceptualizing and therefore mismeasuring the treatment influences our estimates. We could think that only the fact that a city rolls out *a* pop-up bike lane program could already create more cycling. We, therefore, start our investigation by looking at a binary treatment definition, separating the sample into treated and control cities for a standard difference-in-differences analysis. We could, however, also think that the number of kilometers built will make an important difference for the media echo that an announcement of such a policy gets. It is further likely that a kilometer of pop-up bike lane will have a larger impact in a small city than in a larger city. Thus, we also estimate dose-response relationships (effect for each kilometer built) in a generalized difference-in-differences setup and look at effects when the dose is expressed in absolute kilometers and in per capita and per km^2 terms. This helps correct for the fact that in larger or more populous cities, just like in the case of an individual in a cluster RCT in a larger village or school class, we tend to overestimate the dose received at each counter.

All these different cases are presented in Figure 3.3. We show that our estimates get attenuated by measurement error when the dose is expressed in absolute kilometers and that treatment effects are higher, when the dose is expressed relative to population size and relative to the area of a city. Note that our fixed effect already removes any variation between cities in terms of population and area, ensuring that there is no omitted variable bias in the estimates.

Within cities there will still be counters that are farther away from pop-up bike lanes than others and some will only measure the expansion of cycling from pop-up bike lanes partly or not at all. We argue that our fixed effects remove any variation between counters and cities that could lead to systematic relationships between this measurement error and our treatment. With the counter fixed effect we control for factors related to city size, road network, and topography that could be both determinants of counter placement and treatment. The counter fixed effect also controls for local institutions and political majorities that could be driving both where there are counters and where treatment

happens. In our remaining variation, we expect some counters to get a treatment measured with an upward bias and some with a downward bias without there being a systematic tendency, i.e. we are left with classical measurement error.

The counter fixed effect also removes systematic measurement differences between counters, for instance when a counter is placed near a route used for recreational cycling. Further, it ensures that only variation from counters that are present both before and after treatment will be used for the estimation of our treatment effect.

A remaining concern could be that cyclists change their routes and that this could potentially even imply that there is a problem with double-counting. However, we do not expect route changes to create a skewed measurement of cycling traffic. It is likely that cyclists will change their routes in reaction to pop-up bike lanes and this route change means they can come past a counter that they did not pass on their old routes. However, it can also be the case that they switch their route choice away from a counter. On balance, we do not expect this to create systematic measurement error. Further, because our unit of analysis is the counter and we look at average changes in these counts rather than the sum of cyclist counts in a city, there cannot be any double-counting. We also show robustness checks, for which we take the mean of all counters in a city to obtain a measure for city-level cycling traffic and run a regression at the city level (see Figure 3.3 in the paper).

Empirical strategy (continued from main body)

This section provides additional elements regarding our empirical strategy.

Our preferred specifications are presented in Equation 3.1 and Figure 3.2 (marked in blue) in the main body. In Table 3.6 we report on these specifications in more detail, varying the treatment definition (km, km per capita, and km per km^2) and including or dropping the public transit control.

In the following section we discuss our choice of using the outcome (cycling counts) in the logarithmic form. In the next section we explain the additional empirical specification on which Figure 3.2 in the main body is based.

Functional form We use the natural logarithm rather than the level of the count of cyclists as the outcome because we expect cycling to grow in a multiplicative way between cities but also between counters. The country-day FE ensures that we focus on variation between cities in the same country, which are typically at similar stages in their "market penetration" of cycling and where thus cycling can be assumed to grow at similar "natural" rates in the absence of treatment. We expect pop-up bike lanes to have a multiplicative effect across counters in a city because they typically remove prominent bottlenecks from

the network, leading to improved routes and increased (perceived) safety. The cycling counters that measure our outcome are placed next to roads and bike lanes that have different roles in the overall network. Central counters will pick up larger *absolute increases* than more peripheral ones. However, the *growth* rates they measure (approximated by the Δ of the natural logarithms) will be more similar.

We also show that our results are robust in a non-parametric setting using the *Matrix Completion* method for panel data (Athey et al. 2018), a machine-learning method to construct a counterfactual. The Matrix Completion approach frames the problem of causal inference as a missing data problem: For treated units we observe the potential outcome Y^1, i.e. cycling given the introduction of pop-up infrastructure, but we do not observe the potential outcome Y^0 representing cycling in the treated city had the treatment not happened. If we did, the difference would be the treatment effect. We treat our panel dataset as a matrix with missing values, which are the missing potential outcomes Y_i^0. The Matrix Completion method imputes these missing values via regularization based prediction (Athey et al. 2018).

Let Z^0 be the estimated missing elements of the Y^0 matrix. Analytically, the objective of the Matrix Completion approach is to "optimally predict the missing elements by minimizing a convex function of the difference between the observed matrix of Y^0 and the unknown complete matrix Z^0 by using nuclear norm regularization" (Cunningham 2021):

$$\widehat{Z^0} = \arg\min_{Z^0} \sum_{(i,j)\in\Omega} \frac{\left(Y_{it}^0 - Z_{it}^0\right)^2}{|\Omega|} + \Lambda ||Z^0| \tag{3.2}$$

where $||Z^0||$ is the "nuclear norm (sum of singular values of Z^0)" (Cunningham 2021) and Ω denotes the rows i and columns j of the non-missing entries. The regularization parameter Λ is chosen using 10-fold cross-validation.[2]

The coefficients shown in Figure 3.7 confirm that the treatment effect of bike lanes builds up fast enough for specifications with city-week fixed effects to capture it (Equation 3.1, main body).

Additional empirical specification (Figure 3.2 in main body) Estimates shown in Figure 3.2 (main body) are based on the following model:

$$\ln \text{Count}_{id} = \Sigma_\tau \delta_\tau \left(\text{Treated}_c \times D_m^\tau\right) + \mathbf{X}_{cd} + \mu_c + \varphi_{nd} + \varepsilon_{id} \tag{3.3}$$

where i indexes a counter, c a city, n a country, d a day, and m a month.

[2]For a longer explanation of the method in an applied context see Cunningham et al. (2019), who we follow closely here.

The data varies at the counter-day. μ_c is a city fixed effect and φ_{nd} is a country-day fixed effect that captures any daily changes common to all cities in a country.

The coefficients of interest plotted in Figure 3.2 are the δ_τ. They capture the effect of the pop-up bike lane treatment on bicycle counts over time. For this purpose our treatment variable $Treated$ is defined as a binary indicator for treatment that is 1 for treated cities and 0 for control cities. The D_m^τ are binary indicators that are 1 if month m is τ months before or after March 2020, when the pre-treatment period ends. In these specifications, the reference month and begin of the sample is February 2019, when τ equals -13.

Figure 3.2 and Figure 3.8 present the transformed estimate: $100 \times (\exp \delta_\tau - 1)$.

We also run this specification including weather controls $\mathbf{X}_{cd}$ to investigate, if seasonality could be driving these results (Figure 3.8). The results look virtually the same.

SI Appendix Tables

Table 3.2: Summary statistics at the counter-day level

	Mean	Std. Dev.	25%	50%	75%	95%	Min.	Max.
Daily number of cyclists	1457.2	1895.7	255	744	1923	5151	1	13339
City size (ha)	32893.8	42393.6	14163.3	22018.5	40659.9	89180.2	455.6	251517
Year	2017	2	2016	2018	2019	2020	2007	2020
Active counters in city	22.9	23.2	4	14	32	82	1	90
Facebook mobility index	-0.16	0.21	-0.27	-0.11	-0.0044	0.088	-0.81	0.51
Observations	995818							

Notes. The unit of observation of our analysis is the counter and data varies daily. Count data is from municipal bike counters and is obtained from different municipal APIs. Treatment and control variables are assigned to counters based on their city attribute. City definitions are from the EU Urban Audit (Eurostat 2020c). The Facebook mobility index is only available from March 2020. It measures aggregate movement activity by Facebook users in a given administrative area (districts or states).

Table 3.3: Summary statistics of most recent state of infrastructure at the city level

	Mean	Std. Dev.	25%	50%	75%	95%	Max.
Total length of bike infrastructures	11.5	20.0	1.39	2.57	16.6	57.9	85.1
Total length of bike lanes	8.24	18.3	0.24	2.05	7.35	24.8	84.3
Number of measures	19.8	48.1	1	4	17	52	226
Observations	22						

Notes. We use data from the European Cyclists' Federation (ECF 2020). The raw data includes information on individual infrastructure projects announced or implemented. We aggregate it to the city-day level using city definitions from the EU Urban Audit. Our analysis includes data up to July 8, 2020. The newest data can be found at: `https://ecf.com/dashboard`

Table 3.4: Different treatment specifications

	Outcome: Cyclist count			
	(1) All km	(2) Bike lane km	(3) Num of measures	(4) Any treatment
Pop-up treatment	0.006**	0.007**	0.004*	0.061*
	(0.003)	(0.003)	(0.002)	(0.036)
City clusters	78	78	78	78
N	59904	59904	59904	59904

Notes. Each column shows the effect of treatment with pop-up infrastructure on a city's cycling count compiled from city APIs. The data on daily pop-up bike lane additions is from the European Cyclists' Federation (ECF 2020). The newest data can be found at: `https://ecf.com/dashboard`. The unit of observation is the cycling counter. Time variation is daily. Coefficients are from Poisson regressions. Column (1) shows the effect of a kilometer of any bike infrastructure, (2) shows the effect of a kilometer of bike lanes, (3) the effect of any single measure in a city, and (4) the overall treatment of an implemented pop-up infrastructure program in a city. All regressions include counter and day fixed effects and controls for overall mobility (measured with Facebook user movements) (Facebook 2020), weather (temperature, wind, sunshine, precipitation) (Hersbach et al. 2020), and the number of counters active on a given day in a city. We cluster standard errors (in parentheses) at the city level, where treatment is assigned. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Additional robustness checks for our preferred estimate

	(1) Preferred	(2) Transit sample	(3) Large sample	(4) Transit sample w/o Paris	(5) Large sample w/o Paris
Pop-up treatment	30.183***	30.170***	11.197**	34.079***	15.582***
	(8.855)	(8.895)	(5.610)	(6.729)	(3.330)
N	29596	29596	63617	28719	62726

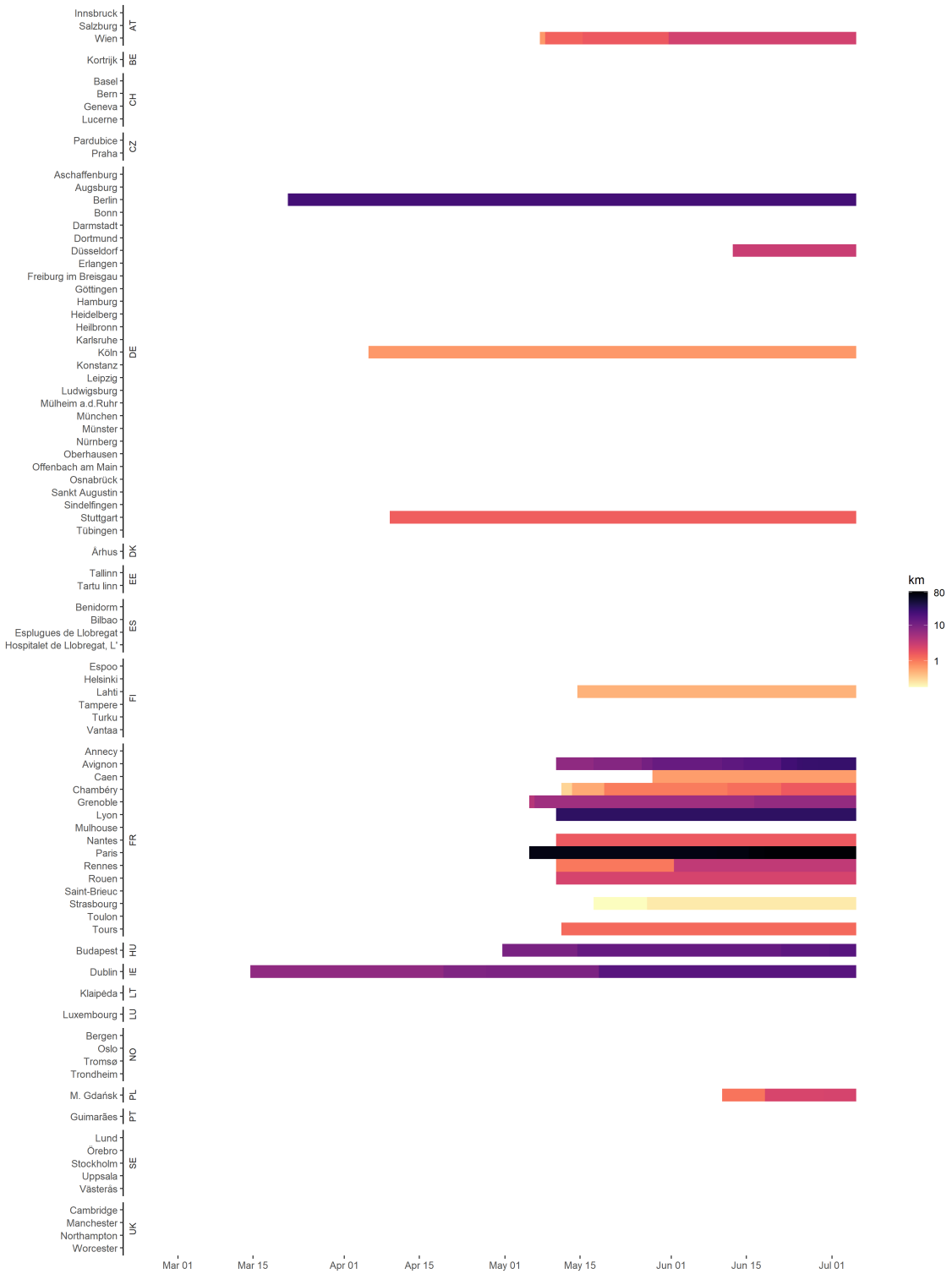
Notes. Each column shows the effect of treatment with pop-up infrastructure on a city's cycling count compiled from city APIs. The data on daily pop-up bike lane additions is from the European Cyclists' Federation (ECF 2020). The newest data can be found at: `https://ecf.com/dashboard`. The unit of observation is the cycling counter. Time variation is daily. Coefficients are from Poisson regressions. Column (1) shows our baseline estimate including the transit control, (2) shows an estimate in the same sample as (1) but without the transit control, (3) the effect in the large sample including cities for which the Apple transit variable does not exist, (4) the same specification as (2) but dropping Paris from the sample, and (5) the same specification as (3) but without Paris. All regressions include counter, city-week, and country-day fixed effects. They also include controls for overall mobility (measured with Facebook user movements) (Facebook 2020), weather (temperature, wind, sunshine, precipitation) (Hersbach et al. 2020), and the number of counters active on a given day in a city. We cluster standard errors (in parentheses) at the city level, where treatment is assigned. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6: Estimates of the average effect of pop-up bike lanes on cycling

	(1)	(2)	(3)	(4)	(5)	(6)
Pop-up bike lanes (km)	7.585***	7.595***				
	(2.545)	(2.538)				
Pop-up bike lanes (km per km^2)			30.183***	30.170***		
			(8.855)	(8.895)		
Pop-up bike lanes (km per capita)					23.295***	23.362***
					(6.124)	(6.148)
Temperature	0.229***	0.231***	0.229***	0.231***	0.229***	0.230***
	(0.048)	(0.049)	(0.048)	(0.049)	(0.048)	(0.049)
Precipitation	-0.035***	-0.036***	-0.035***	-0.036***	-0.035***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
UV Radiation	0.158***	0.158***	0.159***	0.159***	0.159***	0.159***
	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)
Wind (u component)	−0.010	−0.010	−0.011	−0.010	−0.011	−0.011
	(0.009)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
Wind (v component)	−0.008	−0.008	−0.008	−0.008	−0.007	−0.008
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Active Counters (day)	−0.009	−0.012	−0.010	−0.012	−0.010	−0.012
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Overall mobility	1.081***	1.114***	1.076***	1.109***	1.073***	1.105***
	(0.108)	(0.106)	(0.105)	(0.102)	(0.104)	(0.100)
Transit	0.001		0.001		0.001	
	(0.001)		(0.001)		(0.001)	
Counter FE	Y	Y	Y	Y	Y	Y
City-week FE	Y	Y	Y	Y	Y	Y
Country-day FE	Y	Y	Y	Y	Y	Y
City clusters	18	18	18	18	17	17
N	29,596	29,596	29,596	29,596	29,210	29,210
Pseudo R^2	0.932	0.932	0.932	0.932	0.932	0.932

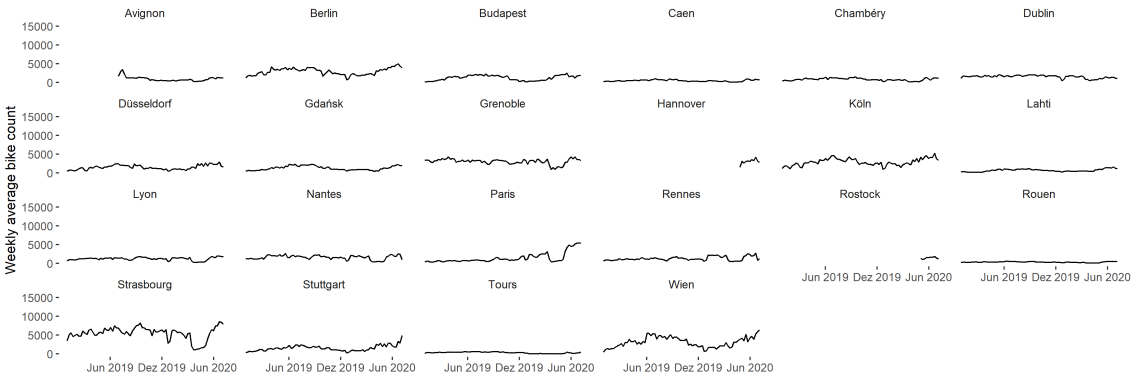
Notes. Estimates are from the preferred specifications marked in blue in Figure 3.3 in the main body (Equation 3.1). These are Poisson regressions using the level of the cyclist count. The unit of observation is the bike counter and data varies daily. Treatment is defined in kilometers, km per capita, or km per km^2 of pop-up infrastructure in service in a city on a day. Treatment effects are scaled to the mean treatment intensity in the sample. Data for the treatment is from the European Cyclists' Federation (ECF 2020) and data for the outcome is from municipal bike counters (Materials and Methods). All regressions include controls for the number of active counters in a city on a given day and for the weather (temperature, sunshine, wind, precipitation; all standardized) (Hersbach et al. 2020). All regressions include a control for overall mobility (Facebook 2020). The transit control is from Apple routing requests (Apple 2020). All regressions include fixed effects at the counter, city-week, and country-day level. We cluster standard errors (in parentheses) at the city level, where treatment is assigned. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.4: Intensity of pop-up bike lane treatment over time in treatment cities and control cities



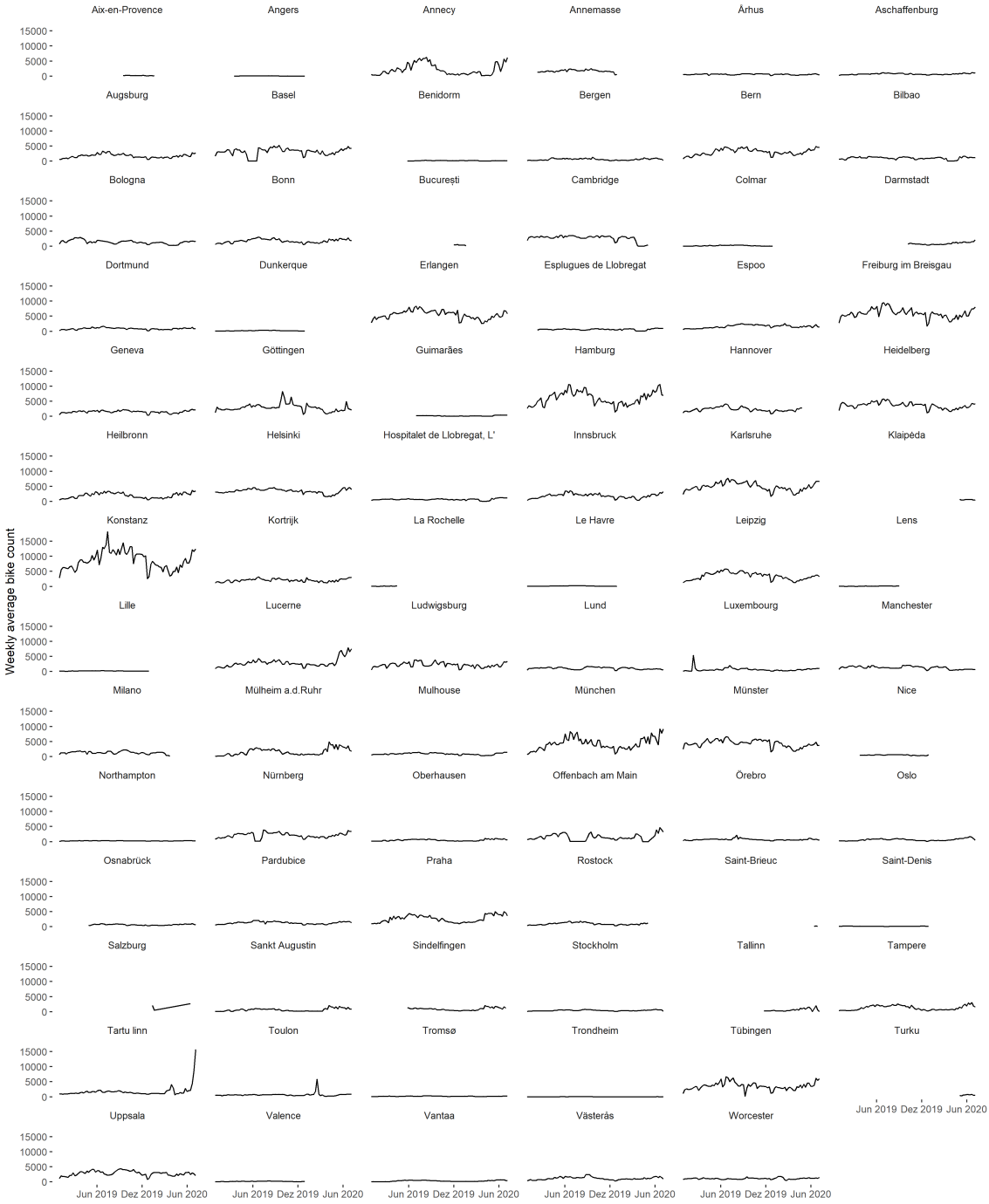
Notes. This figure shows treated cities and their treatment intensities in terms of kilometers (coloring on a log scale) of public bike lanes in service on a given day between March and July 2020. Note that some control cities have implemented pop-up bike lanes after July 2020. London, Milan, Lisbon and Rome are missing from the sample due to insufficient spatial or temporal coverage of the data. Information on individual pop-up bike lanes with their street location, announcement date, and implementation status is from the European Cyclists' Federation (ECF 2020). The newest data can be found at: `https://ecf.com/dashboard`

Figure 3.5: Average bike count per week in treated cities



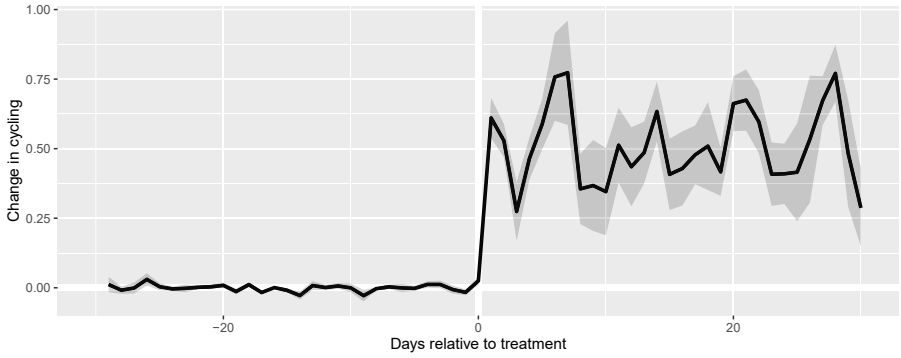
Notes. Daily bike counts are aggregated by city and averaged over the week. Bike counts are assembled from municipal open data feeds. The lower and upper percentiles of the base sample (treated and control cities combined) are removed from the estimation sample. Only measurements from 2019 and 2020 are shown. City definitions are chosen according to the EU Urban Audit (Eurostat 2020c).

Figure 3.6: Average bike count per week in control cities



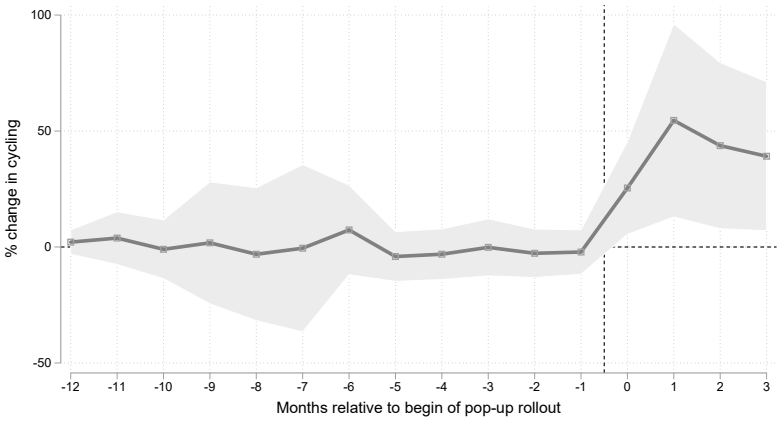
Notes. Daily bike counts are aggregated by city and averaged over the week. Bike counts are assembled from municipal open data feeds. The lower and upper percentiles of the base sample (treated and control cities combined) are removed from the estimation sample. Only measurements from 2019 and 2020 are shown. City definitions are chosen according to the EU Urban Audit (Eurostat 2020c).

Figure 3.7: Change in daily cycling after the first day of treatment for each city (see Equation 3.2)



Notes. The figure shows the average treatment effect on the treated (ATT) of treatment with pop-up bike lanes based on the *Matrix Completion* method (Athey et al. 2018). We implement Matrix Completion at the counter-day level and include controls for the number of active counters in a city on a given day and for the weather (temperature, sunshine, wind, precipitation) (Hersbach et al. 2020). The lighter grey area shows 95% confidence intervals based on 5000 bootstrap runs clustered at the city level. The sample for this Figure is restricted to 30 days before and after the first treatment day for each counter/city. Note that estimates are not converted to % changes. The figure is implemented with the *gsynth* package (Xu 2017).

Figure 3.8: Treatment effect (difference between treated and control cities) in months before and after the beginning of the pop-up bike lane policy



Notes. This is the same specification as shown in Figure 3.2 in the main body (Equation 3.3) except for the inclusion of weather controls (Hersbach et al. 2020). Observations are binned into months. Treatment for this plot is hard-coded to March 2020 and the baseline category and the begin of the sample are set to February 2019. Estimates are from Poisson regressions that include city and country-day fixed effects. The shaded area shows the 95% confidence interval. Data for the outcome variable is from the European Cyclists' Federation (ECF 2020) and data for the treatment variable is from municipal bike counters (Materials and Methods).

3.6 References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge (2017). *When Should You Adjust Standard Errors for Clustering?* Working Paper 24003. National Bureau of Economic Research. DOI: 10.3386/w24003.

Agyeman, Julian (2020). *Poor and Black 'invisible Cyclists' Need to Be Part of Post-Pandemic Transport Planning Too*. http://theconversation.com/poor-and-black-invisible-cyclists-need-to-be-part-of-post-pandemic-transport-planning-too-139145.

Aldred, Rachel (2019). "Built Environment Interventions to Increase Active Travel: A Critical Review and Discussion". *Current Environmental Health Reports* 6.4, pp. 309–315. DOI: 10.1007/s40572-019-00254-4.

Aldred, Rachel, Bridget Elliott, James Woodcock, and Anna Goodman (2017). "Cycling Provision Separated from Motor Traffic: A Systematic Review Exploring Whether Stated Preferences Vary by Gender and Age". *Transport Reviews* 37.1, pp. 29–55. DOI: 10.1080/01441647.2016.1200156.

Apple (2020). *Mobility Trends Reports*. https://www.apple.com/covid19/mobility.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi (2018). *Matrix Completion Methods for Causal Panel Data Models*. Working Paper 25132. National Bureau of Economic Research. DOI: 10.3386/w25132.

Bezirksamt Friedrichshain-Kreuzberg (2020). *Einrichtung von pandemieresilienter Infrastruktur in Form von temporären Radverkehrsanlagen*. https://www.berlin.de/ba-friedrichshain-kreuzberg/aktuelles/pressemitteilungen/2020/pressemitteilung.937004.php.

Buckee, Caroline O., Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H. Grad, Bryan Grenfell, M. Elizabeth Halloran, Moritz U. G. Kraemer, Marc Lipsitch, C. Jessica E. Metcalf, Lauren Ancel Meyers, T. Alex Perkins, Mauricio Santillana, Samuel V. Scarpino, Cecile Viboud, Amy Wesolowski, and Andrew Schroeder (2020). "Aggregated Mobility Data Could Help Fight COVID-19". *Science* 368.6487, pp. 145–146. DOI: 10.1126/science.abb8021.

Buehler, Ralph and Jennifer Dill (2016). "Bikeway Networks: A Review of Effects on Cycling". *Transport Reviews* 36.1, pp. 9–27. DOI: 10.1080/01441647.2015.1069908.

Buehler, Ralph and John Pucher (2012). "Cycling to Work in 90 Large American Cities: New Evidence on the Role of Bike Paths and Lanes". *Transportation* 39.2, pp. 409–432. DOI: 10.1007/s11116-011-9355-8.

Buehler, Ralph, John Pucher, and Adrian Bauman (2020). "Physical Activity from Walking and Cycling for Daily Travel in the United States, 2001–2017: Demographic, Socioeconomic, and Geographic Variation". *Journal of Transport & Health* 16, p. 100811. DOI: 10.1016/j.jth.2019.100811.

Chang, Hung-Hao, Chad Meyerhoefer, and Feng-An Yang (2020). *COVID-19 Prevention and Air Pollution in the Absence of a Lockdown*. Working Paper 27604. National Bureau of Economic Research. DOI: 10.3386/w27604.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin (2020). "Fast Poisson Estimation with High-Dimensional Fixed Effects". *The Stata Journal* 20.1, pp. 95–115. DOI: 10.1177/1536867X20909691.

Cunningham, Scott (2021). *Causal Inference: The Mixtape*. Yale University Press.

Cunningham, Scott, Gregory DeAngelo, and John Tripp (2019). *Craigslist Reduced Violence Against Women*. Working Paper.

Dill, Jennifer and Theresa Carr (2003). "Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them". *Transportation Research Record* 1828.1, pp. 116–123. DOI: 10.3141/1828-14.

Dill, Jennifer and Nathan McNeil (2016). "Revisiting the Four Types of Cyclists: Findings from a National Survey". *Transportation Research Record* 2587.1, pp. 90–99. DOI: 10.3141/2587-11.

ECF (2020). *COVID-19 Cycling Measures Tracker*. https://ecf.com/dashboard. Brussels.

Eurostat (2020a). *Geostat grid*.

– (2020b). *Transport - Cities and Greater Cities (Urb_ctran)*.

– (2020c). *Urban Audit Geodata (GISCO)*.

Facebook (2020). *Movement Range Maps*. https://data.humdata.org/dataset/movement-range-maps.

Hamilton, Timothy L. and Casey J. Wichman (2018). "Bicycle Infrastructure and Traffic Congestion: Evidence from DC's Capital Bikeshare". *Journal of Environmental Economics and Management* 87, pp. 72–93. DOI: 10.1016/j.jeem.2017.03.007.

Healy, Kieran (2020). *covmobility: Mobility Data from Apple and Google*. R package version 0.1.0.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut (2020). "The ERA5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049. DOI: 10.1002/qj.3803.

Hubrich, Stefan, Frank Ließke, Rico Wittwer, Sebastian Wittig, and Regine Gerike (2019). *Methodenbericht Zum Forschungsprojekt - Mobilität in Städten–SrV 2018*.

Hufkens, Koen, Reto Stauffer, and Elio Campitelli (2019). *The ecmwfr package: an interface to ECMWF API endpoints*. https://bluegreen-labs.github.io/ecmwfr/. DOI: 10.5281/zenodo.2647541.

Kraus, Sebastian and Nicolas Koch (2020). *Effect of Pop-up Bike Lanes on Cycling in European Cities (Code and Data)*. 10.5281/zenodo.4015974. Zenodo.

Larcom, Shaun, Ferdinand Rauch, and Tim Willems (2017). "The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network". *The Quarterly Journal of Economics* 132.4, pp. 2019–2055. DOI: 10.1093/qje/qjx020.

Manaugh, Kevin, Geneviève Boisjoly, and Ahmed El-Geneidy (2017). "Overcoming Barriers to Cycling: Understanding Frequency of Cycling in a University Setting and the Factors Preventing Commuters from Cycling on a Regular Basis". *Transportation* 44.4, pp. 871–884. DOI: 10.1007/s11116-016-9682-x.

Marqués, R., V. Hernández-Herrador, M. Calvo-Salazar, and J. A. García-Cebrián (2015). "How Infrastructure Can Promote Cycling in Cities: Lessons from Seville". *Research in Transportation Economics*. Bicycles and Cycleways 53, pp. 31–44. DOI: 10.1016/j.retrec.2015.10.017.

Mattauch, Linus, Monica Ridgway, and Felix Creutzig (2016). "Happy or Liberal? Making Sense of Behavior in Transport Policy Design". *Transportation Research Part D: Transport and Environment* 45, pp. 64–83. DOI: 10.1016/j.trd.2015.08.006.

McNeil, N., J. Broach, and Jennifer L. Dill (2018). "Breaking Barriers to Bike Share: Lessons on Bike Share Equity". *Ite Journal-institute of Transportation Engineers* 88.

Mueller, Natalie, David Rojas-Rueda, Maëlle Salmon, David Martinez, Albert Ambros, Christian Brand, Audrey de Nazelle, Evi Dons, Mailin Gaupp-Berghausen, Regine Gerike, Thomas Götschi, Francesco Iacorossi, Luc Int Panis, Sonja Kahlmeier, Elisabeth Raser, and Mark Nieuwenhuijsen (2018). "Health Impact Assessment of Cycling Network Expansions in European Cities". *Preventive Medicine* 109, pp. 62–70. DOI: 10.1016/j.ypmed.2017.12.011.

Ortiz-Bobea, Ariel (2020). *ArielOrtizBobea/Spec_chart*. https://github.com/ArielOrtizBobea/spec_chart.

Parkin, John, Mark Wardman, and Matthew Page (2008). "Estimation of the Determinants of Bicycle Mode Share for the Journey to Work Using Census Data". *Transportation* 35, pp. 93–109. DOI: 10.1007/s11116-007-9137-5.

Pucher, John, Jennifer Dill, and Susan Handy (2010). "Infrastructure, Programs, and Policies to Increase Bicycling: An International Review". *Preventive Medicine* 50, S106–S125. DOI: 10.1016/j.ypmed.2009.07.028.

Salmon, Maëlle and Natalie Mueller (2017). *masalmon/cycle_infrastructure_modeshare: Zenodo first version*. 10.5281/zenodo.322906.

Schoner, Jessica E. and David M. Levinson (2014). "The Missing Link: Bicycle Infrastructure Networks and Ridership in 74 US Cities". *Transportation* 41.6, pp. 1187–1204. DOI: 10.1007/s11116-014-9538-1.

Winters, Meghan, Ralph Buehler, and Thomas Götschi (2017). "Policies to Promote Active Travel: Evidence from Reviews of the Literature". *Current Environmental Health Reports* 4.3, pp. 278–285. DOI: 10.1007/s40572-017-0148-x.

Xu, Yiqing (2017). "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models". *Political Analysis* 25.1, pp. 57–76. DOI: 10.1017/pan.2016.2.

Yang, Lin, Shannon Sahlqvist, Alison McMinn, Simon J. Griffin, and David Ogilvie (2010). "Interventions to Promote Cycling: Systematic Review". *BMJ* 341. DOI: 10.1136/bmj.c5293.

Zapata-Diomedi, Belen, Lucy Gunn, Billie Giles-Corti, Alan Shiell, and J. Lennert Veerman (2018). "A Method for the Inclusion of Physical Activity-Related Health Benefits in Cost-Benefit Analysis of Built Environment Initiatives". *Preventive Medicine* 106, pp. 224–230. DOI: 10.1016/j.ypmed.2017.11.009.

Chapter 4

No aggregate deforestation reductions from roll-out of community land titles in Indonesia yet[†,‡]

Abstract

In Indonesia, 60 million people live within 1 kilometer of state forest. The government of Indonesia plans to grant community titles for 12.7 million hectares of land to communities living in and around forests. These titles allow for using non-timber forest products, practicing agro-forestry, operating tourism businesses, and selective logging in designated production zones. Here, we estimate the early effects of the program's roll-out. We use data on the delineation and introduction date of community forest titles on 2.4 million hectares of land across the country. We find that, contrary to the objective of the program, community titles aimed at conservation did not decrease deforestation; if anything, they tended to increase forest loss. In contrast, community titles in zones aimed at timber production decreased deforestation, albeit from higher baseline forest loss rates.

4.1 Introduction

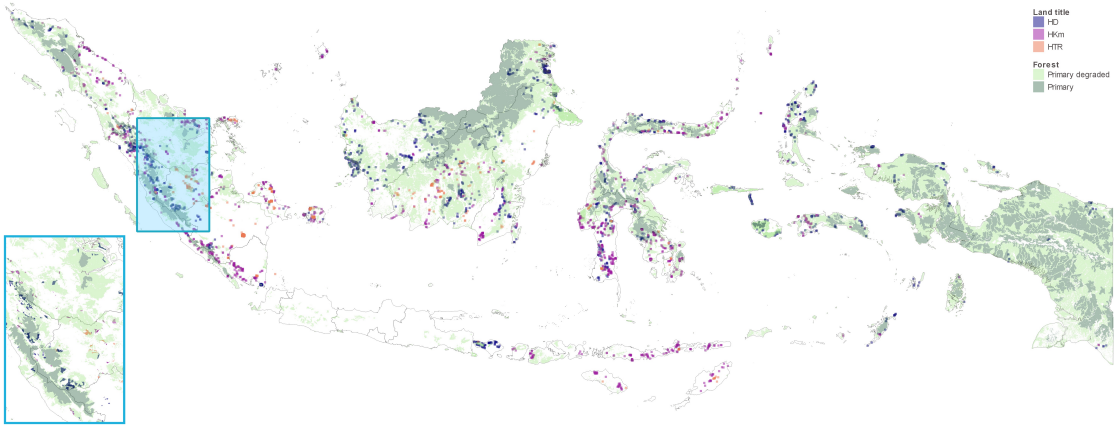
In 2015 Indonesia was the world's fourth largest emitter of greenhouse gas emissions. Almost 60% of its carbon emissions were from land use change (World Resources Institute et al. 2019). After 2016 deforestation in Indonesia has fallen by 30% (Comparing 2009–2016 with 2017–2019) (Hansen et al. 2013). This has in part been credited to a policy mix including bans on primary forest clearing and peat drainage, a review of land concessions, and a moratorium on new palm oil plantations and mines (Busch et al. 2015). Meanwhile, certification has helped protecting forests on existing plantations (Carlson et al. 2018). Indonesia has also initiated a large-scale land titling program, that consists of the release of land for agriculture and a "social forestry" component. The social forestry program aims to title 12.7 million hectares as community forest. By improving livelihoods, resolving tenure conflicts, and involving communities in forest management, the program also aims at slowing deforestation. Here, we investigate the early effects of the roll-out of the Indonesian social forestry policy on forest loss. We analyze new data on the extent and the titling year of 4,349 land titles covering 2.4 million hectares (median size 70 hectares). Our sample of titled areas runs from 2009 to 2019 and titling accelerated markedly after 2016. We compare these treated areas to control areas (median size 55 hectares) from the pool of candidate areas designated for titling by the Ministry of Environment (*PIAPS* map). We use satellite-based measurements of annual forest loss between 2001 and 2019

[†]Sebastian Kraus, Jacqueline Liu, Nicolas Koch, and Sabine Fuss (2021). "No Aggregate Deforestation Reductions from Rollout of Community Land Titles in Indonesia Yet". *Proceedings of the National Academy of Sciences* 118.43. DOI: 10.1073/pnas.2100741118

[‡]We thank Mhabeni Bhona for background research and support with research design.

(Hansen et al. 2013; Margono et al. 2014) to compare changes in treatment and control areas before and after the introduction of each land title with a regression-based stacked difference-in-differences design.

Figure 4.1: Community title area locations and primary forest



Notes. This map shows the three main types of social forestry HD and HKm, which allow non-timber forest product collection, agroforestry, ecotourism, and some selective logging, and HTR, which aims at restoring degraded areas for community timber plantations. Primary forest and primary degraded forest in 2000 are shown in the background (Margono et al. 2014) and province boundaries in gray.

In Indonesia much of the land on the outer islands is designated as state forest zone. Consequently, 6 million people live on and 60 million within one kilometer of land designated as state forest (own estimates, SI Appendix, Section 4.5). The population close to forest frontiers still relies at least partially on land for their livelihoods, either for non-timber forest products and smallholder agriculture or for ownership of or work on plantations. Community titles have been described as a tool to foster the sustainable use of natural resources, but governments tend to use them in areas with low pressure on forests. It has been unclear, whether they would work at deforestation frontiers. In contrast with typical programs in other countries, the Indonesian social forestry program targets comparatively densely populated, fragmented landscapes with high pressure on forest. Many of the land titles are granted for areas on the edges of primary forest (Fig. 4.1).

The social forestry program's theory of change is that giving communities land use rights can improve environmental outcomes since both land conflicts and poverty are important drivers of forest loss that community titles help address (The World Bank 2020). A community title may lead to increased conservation if village institutions are designed to foster coordination and enforcement (Ostrom 1990) or if tenure security lowers the risk related to punishment for agroforestry practices and the use of non-timber forest products, discourages short-term resource exploitation, or makes credit available for the enhancement of forest resources. Under favorable local conditions these mechanisms can lead to more sustainable forms of consumption-smoothing, investments in restoration, or intensification outside of forested land (Angelsen and Kaimowitz 1999). However, in

theory, the introduction of a land title can also lead to additional land clearing if a reliance on community institutions weakens overall governance (Ostrom 1990) or if market forces favor an expansion of plantation-based cash crops rather than forest-based cash crop production or improved subsistence agriculture (Angelsen and Kaimowitz 1999). In line with these inconclusive theoretical predictions, previous empirical research has found mixed effects of land tenure programs on deforestation (Börner et al. 2020; Tseng et al. 2020) (SI Appendix, Section 4.5).

4.2 Results

We estimate the average country-wide effect of the community titling policy and test one of its main hypotheses: That community land titling reduces forest loss. Indonesian social forestry titles allow communities to use non-timber forest products, practice agro-forestry, and operate eco-tourism businesses in areas designated as protection zones by the government. In government-designated production zones, they also allow selective logging and, in specific cases (HTR titles), timber plantations.

We differentiate between the three main types of land titles: (i) village forests (*hutan desa*, HD), (ii) community forests (*hutan kemasyarakatan*, HKm), and (iii) community plantation forest (*hutan tanaman rakyat*, HTR). These titles are granted to villages (HD), cooperatives (HKm and HTR), or individual farmers (HTR) for 35 years (Rakatama and Pandit 2020). HD and HKm allow for restricted (50 m^3 per year) logging for non-commercial purposes and conditional on avoiding "net deforestation" (Santika et al. 2019) and only in areas defined as production zones by the government. For HD titles this is the case for 42% of the total titled area. HTR titles are established on plantation forests and degraded areas, for which communities or farmers can obtain licenses to operate and restore timber plantations. Our main sample contains 950 HD areas with a median size of 807.8 hectares, 1224 HKm areas with a median size of 246 hectares, and 2175 HTR areas with a median size of 1.3 hectares.

We compared areas with land titles before and after approval to control areas, which are planned to get community titles in the future, in a stacked difference-in-differences design using Poisson regressions. The estimates in Table 4.1 can be interpreted as changes in deforestation rates in treated areas compared to control areas.

We find that the main types of community titles (HD and HKm) on average do not decrease deforestation rates (columns (A) to (F) in Table 4.1). If anything, we find increases in forest loss for these two pillars of the social forestry program. However, the 95% confidence intervals for the positive estimates in most cases include increases in deforestation that are small or even slightly negative. For HD titles we can thus rule out substantial reductions in deforestation for for all types of tree cover pooled into one category (A), degraded primary forest (B), and undisturbed primary forest (C). This result contrasts with earlier

findings based on a sample of 93 HD areas (Santika et al. 2017). Our results are stable to specifications with sample restrictions in terms of included cohorts, types of forest zones, and forest loss year observations that make our estimation sample similar to the earlier study (Santika et al. 2017) based on a smaller sample. This difference could result from selection into treatment effects because titles may be granted for the most promising areas first. It could also point to remaining omitted variable bias in the earlier study, which our stacked difference-in-difference design helps to correct (Materials and Methods).

An exception to the overall lack of reductions in deforestation rates can be found for HTR areas, which are granted on plantation forests. HTR areas have higher forest loss rates than HD and HKm areas before the introduction of the community title (around double for all tree cover combined and 50% more for degraded primary forest areas). For these titles, we find substantial decreases in forest loss rates on degraded primary forest (column (H) in Table 4.1). This is indicative of increased efforts to restore forests for timber production in the HTR sub-sample.

4.3 Discussion

We find that the main building blocks of the Indonesian community titling program (HD and HKm titles) overall have not led to decreases in forest loss. Case-study evidence points to two main explanations for this result: (i) a lack of institutional capacity at the community level (Rakatama and Pandit 2020) and (ii) the economic opportunity costs of conservation (Langston et al. 2017). Many communities lack the resources to monitor their areas or agree upon and enforce rules on resource use (Rakatama and Pandit 2020). In addition, in contexts with suitable market-access for plantation-based cash crops, a reduced risk of expropriation may have led to investments in land-clearing rather than restoration or forest-based activities (Langston et al. 2017). The main pillars of the program (HD and HKm) allow only the extraction of non-timber forest products and activities based on eco-system services (for instance tourism) or selective logging depending on government zoning. These activities may not provide sufficient incentives to increase conservation efforts. Payments for ecosystem services, for instance based on ecological intergovernmental transfers to villages, could help increase the value of conservation for these communities above the opportunity costs of logging and agriculture (Wunder et al. 2020) or help subsistence households meet consumption needs and reduce their reliance on forests as a safety-net (Ferraro and Simorangkir 2020).

Prior research has found forest loss reductions in a sample of early social forestry areas (Santika et al. 2017). This result indicates that under favorable conditions the policy can reduce deforestation. Our result indicates that on aggregate the program does not deliver these reductions yet.

For a sub-part of the program focused on community timber production on plantations

and in degraded areas (HTR titles), we find evidence for forest loss reductions indicating an opportunity for increased conservation by including Indonesian communities in efforts to restore degraded plantations. However, HTR titles only constitute 6.3% of the granted social forestry area and their median size is small (1.3 ha).

Table 4.1: Effect of community land titling on deforestation

	HD			HKm			HTR	
	(A) All forest	(B) Degraded	(C) Primary	(D) All	(E) Degraded	(F) Primary	(G) All	(H) Degraded
Land title	0.14**	0.27***	0.35*	0.08	0.00	0.75**	-0.10	-0.83***
	[0.02,0.26]	[0.08,0.46]	[-0.03,0.73]	[-0.06,0.21]	[-0.26,0.27]	[0.16,1.34]	[-0.24,0.05]	[-1.21,-0.46]
Precipitation	-0.15***	-0.16***	-0.32***	-0.15***	-0.16***	-0.31***	-0.15***	-0.16***
	[-0.20,-0.11]	[-0.23,-0.09]	[-0.52,-0.11]	[-0.20,-0.11]	[-0.23,-0.09]	[-0.51,-0.11]	[-0.20,-0.11]	[-0.22,-0.09]
Clusters	18552	10554	1497	18746	10438	1416	19259	10246
N	3714021	2073305	287727	3717707	2071101	286188	3727454	2067453

Notes. Estimates from Poisson regressions on a stacked sample of treatment and control groups. The row *Land title* reports the coefficient on the interaction term between an indicator for treatment and an indicator for years after treatment (Equation 4.1, SI Appendix, Section 4.5). The unit of analysis is the study area. Treated units are areas with community titles and control units are areas designated for treatment by the government. Standard errors are clustered at the study area level, where treatment is assigned (see number of *Clusters*). The number of units of analysis corresponds to the number of clusters. The total number of observations (N) corresponds to all units and years in the stacked panel. We differentiate between the three types of social forestry HD, HKm, and HTR. The outcome is the deforestation rate, i.e. area deforested divided by total area, at the level of the unit of observation. We show results for deforestation rates in all forest combined (Hansen et al. 2013) and restricted to degraded primary forest and primary forest (Margono et al. 2014). All regressions include study area fixed effects, year fixed effects, a fixed effect indicating, if an observation is before or after the treatment year of a cohort, and a fixed effect indicating, if an observation is in the treated or in the control group for a given cohort. All regressions control for annual precipitation (CHIRPS, standardized). The 95% confidence interval is shown in brackets. Significance levels are indicated by *, **, and *** for 10%. 5%, and 1% respectively.

4.4 Materials and methods

Land title data

We use boundaries and land title identifiers of community titles as published by the Indonesian Ministry of Environment and Forestry (version: 14 September 2020). We also use areas designated for social forestry by the Ministry, which serve as a control group (*PIAPS* map, SI Appendix, Section 4.5).

Forest data

We use version 1.7 of the Hansen et al. (2013) Global Forest Change data and the Margono et al. (2014) natural forest data for Indonesia. We reclassify the Margono categories of primary forest into primary (2, 4–9, 13–15) and primary degraded forest (1, 10–12). The outcome *all forest* is based on the Hansen data only, which detects any tree cover change including on plantations. We use Google Earth Engine to extract the annual sum of deforested hectares for treatment and control areas (script in code repository). The outcome variable *annual deforestation rate* is the sum of deforested hectares divided by the total size of an area.

Econometrics

We use a stacked difference-in-differences design with Poisson regressions to estimate the effect of a land title on deforestation rates. We leverage variation in the timing of treatment and use areas that have not been titled yet as a counterfactual.

The main empirical challenge is to construct credible counterfactuals for the treated areas. Governments, communities and NGOs may work towards a title for a specific area, if it is particularly easy to lower deforestation there. However, they may also react to high local pressure on forests leading to bias in the opposite direction. Often the underlying factors are time-variant and difficult to measure. Therefore, the direction of the bias in a research design relying on control areas matched based on measurable factors, such as topography or climate, would be theoretically unclear. We compare areas with land titles before and after approval to control areas that serve as counterfactual units. These counterfactual units are areas designated by the Indonesian Ministry of Environment and Forestry to get community titles in the future (*PIAPS* map, SI Appendix, Section 4.5). Thereby we avoid using already treated units as control units for later cohorts, which could otherwise lead to violations of the common trends assumption (Goodman-Bacon 2018).

Data and code

Repository to replicate the analysis and robustness checks: https://zenodo.org/record/4314768

4.5 SI Appendix

Background

Land titling in Indonesia

There are two overarching land categories in Indonesia: forest estate (*Kawasan Hutan*), regulated by forestry law, and area for other land use (*Areal Penggunaan Lain*, APL), regulated by Basic Agrarian Law (UUPA No.5/1960).

Stemming from the legacy of the colonial government, by default, forest in Indonesia has been owned by the state. The Basic Forestry Act of 1967 claimed around 70% of land as state-administered forest. This law was later replaced with Forest Law 41/1999. The new forest law became the legal basis for the social forestry program (Resosudarmo et al. 2019; Siscawati et al. 2017). It states that "forest and land rehabilitation shall be chiefly implemented primarily by [a] participatory approach to develop [the] potential [of] and empower community" (Art. 42, §2). The law defines social forests as "state forest utilized for empowering communities".

As part of the initial wave of decentralization after the fall of Suharto, districts gained the authority to issue small-scale forest concessions to communities and cooperatives (Dermawan et al. 2006). However, many of these rights were retracted not long after their introduction. Dermawan et al. (2006) documents the recentralization measures that the central government introduced from the end of 1999, and the corresponding power struggle between the central government and sub-national forest authorities. Ultimately, however, new legislation on social forestry implementation in 2008 and the Widodo administration's (assumed office in 2014) social forestry targets (12.7 ha) have initiated an accelerated roll-out of the policy.

The forest land titles in our study (HD, HKm, HTR) are granted for land in state forest for a period of 35 years (see Ministry of Environment and Forestry Indonesia (2021a) for an interactive map, albeit without area boundaries). The titling process is coordinated by the Ministry of Environment and Forestry. Based on suggestions from local government and civil society the Ministry publishes and updates maps of candidate areas designated for social forestry (*Peta Indikatif dan Areal Perhutanan Sosial, PIAPS*, Indicative and Social Forestry Map) (Ministry of Environment and Forestry Indonesia 2021b). These areas

are the control areas in our study. The process of designation of candidate areas by the Ministry of Environment and Forestry is part of the government's *One Map* initiative. The initiative's goal is to harmonize land maps within the Ministry and between Ministries and thereby solve overlapping land claims. Until 2016, the administrative procedures for the actual titling of these candidate areas were split between national and local government levels and consisted in a two-step process with an initial certificate and a final permit and a verification process in between (Siscawati et al. 2017). For instance, permits for HD and HKm areas had to pass around 30 desks in more than 4 offices taking half a year (Fisher et al. 2018). Since 2016 the titling process has been streamlined with decisions centralized at the Ministry through its Directorate General of Social Forestry and Environmental Partnerships (Ministerial Decree 83/2016).

Indigenous community titles (*Hutan Adat*), that we do not study, are an exception to the general principle that titled land technically remains in the state forest estate. In 2013, Indonesia's constitutional court ruled that customary forests should no longer be part of state forests (MK35/2012). As of the time of the analysis, only 66 such titles with a total area of around 44,600 hectares had been granted. Therefore these titles have not been included in the analysis. A fifth social forestry scheme consists of partnerships (*kemitraan*) between plantation firms and communities. This scheme is not included in our analysis because of a lack of available data. Note that the roll-out of social forestry titles is accompanied by a push to title agrarian lands (*Tanah Objek Reformasi Agraria*, TORA), for which some land is also "released" from the forest estate. Our analysis does not include these non-forest land titles.

The state forest estate is classified into different land use zones, including production, limited production, conversion, conservation, and watershed protection forest. The land titles we study grant communities additional rights, but they still have to comply with the legal framework created by this general zoning. In contrast with HTR titles, that are aimed at timber plantations, HD and HKm titles restrict timber extraction to selective logging and to areas outside of watershed protection forest (*Hutan Lindung, HL*). In limited production (HPT), production (HP) and conversion (HPK) areas HD and HKm permits allow the extraction of 50 m^3 of timber per year for non-commercial use conditional on avoiding "net deforestation" (Ministerial Decree No. P. 89/2014 Article 33 (Santika et al. 2019)).

Theory

One of the several goals of the social forestry program is to reduce deforestation. The program's theory of change builds on the assumption that titling can create synergies between land rights, livelihoods, and conservation (The World Bank 2020). However, from a theoretical point of view the average, net effect to be expected of the program is unclear (Angelsen and Kaimowitz 1999) as there can be trade-offs between land rights and conservation goals (Hajjar et al. 2020) and between poverty reduction and conservation (Chhatre and Agrawal 2009; Sunderlin et al. 2005). Here, we provide a theoretical framework documenting the mechanisms that can influence deforestation rates in either direction after the introduction of a social forestry permit, which leads us to the conclusion that a reduced-form quantitative evaluation of the whole program can generate valuable insights.

Institutional challenges

The social forestry program relies on institutions at the village or community level that ensure that the community solves the "open access" problem by agreeing on rules regarding the use of their resource and by managing to implement them in practice through monitoring and sanctions. Case studies of social forestry projects in Indonesia report on problems with institutional capacity (Rakatama and Pandit 2020) – in many instances with substantial challenges meeting several of Ostrom's (1990) eight design principles for the sustainable governance of common pool resources. A recent review summarizes challenges to the functioning of social forestry institutions described in the case-study literature (Rakatama and Pandit 2020). First, while social forestry areas create increased clarity on the boundaries of a community's forest, boundaries are not always recognized by external actors and encroachment can persist (challenge to Principle 1). Second, some case studies note that social forestry as a national-level institution only takes into account local informal and customary institutions to a limited extent. The national social forestry policy pre-defines most of the rules and administrative procedures of the institution leaving less room for local rule-making (Principles 2, 3, and 7). Inequalities within villages persist both in terms of access to resources and decision-making (Principle 3). Third, transaction costs created by the administration of the program and the fulfillment of government-set criteria may be too high compared to the benefits of the policy (Principle 2). Fourth, private sector actors or other levels of governance are reluctant to recognize community rights and to transfer power to the new institution (Principles 7 and 8). Fifth, community institutions lack the capacity and authority to monitor their social forestry areas (Principle 4), resolve conflicts (Principle 5), and sanction breaches of the common rules (Principle 6). Finally, note that the government has managed to reduce overall deforestation rates in Indonesia after 2016. It is still unclear, how this reduction was achieved exactly, but anecdotal evidence points to increased law enforcement playing an important role (Wijaya et al. 2019). A partial "outsourcing" of monitoring and law enforcement to communities

on their social forestry areas may lead to higher deforestation rates there compared to control areas without title.

Incentives

However, beside institutional issues we also expect incentives to be an important driver of the decision to conserve or to deforest not least because community members need to dedicate time and resources to social forestry institutions including the monitoring of the areas. Incentives can be related to both non-monetary and monetary elements that provide utility to the community. This may include norms and values, traditions, and personal preferences for conservation. Some of these elements will be partially reflected in economic outcomes, such as through air and water pollution or landslides. From other elements communities will only derive non-monetary utility. On the monetary side, incentives for communities depend fundamentally on their market access for different cash crops, which can range from non-timber forest products over agroforestry-based to plantation-based crops. If cash crop agriculture is profitable in a location, it will be an important determinant of land use decisions beside the subsistence needs, for which forest land can be a regular source of consumption goods (Angelsen et al. 2014) or act like an insurance (Ferraro and Simorangkir 2020) (see also Wunder et al. (2014) for a survey-based review of the "safety-net" function of forests, which quantitatively may be more limited than described in case studies). The social forestry program aims at improving livelihoods through regularizing and providing more long-term security of forest-based economic activities. Oil palm plantations are defined as non-forest based activities and are thus technically excluded even for HTR community plantations. However, land clearing can still be more profitable than conservation-based economic activities even considering the cost of breaching government zoning rules. In the following, we explain how social forestry permits can change the utility of conservation and its opportunity cost.

The introduction of a social forestry permit has three first order effects that influence the utility communities derive from their land. First, it lowers (or removes) the expected costs related to breaches of government zoning. Second, it reduces the uncertainty of future returns related to expropriation risk. Third, a social forestry title may facilitate a community's access to credit. All three mechanisms can theoretically shift the balance between the benefits and the opportunity cost of conservation in either direction. We explain them in more detail in the following paragraphs.

First, social forestry permits change the costs related to government sanctions. Anecdotal evidence points to high costs related to potential punishment before the introduction of permits. These costs can influence both the benefits of conservation but also of land-clearing (Koch et al. 2019). Communities report fearing fines or other forms of sanctions for instance for accessing their agroforestry plantations and thus concentrating their work hours before and after dawn. With the introduction of a social forestry permit for fully legal activities

only the costs of compliance with government regulations remain, for instance regarding the maximum volume of logging or the canopy density for agroforestry plantations. On the other hand, for activities that are not in line with government regulations, such as land clearing for agriculture, social forestry permits might also be perceived as a signal that the government may become less present on the titled lands and thus sanctions would become less likely. A related mechanism may be that the exact rights and duties related to social forestry permits are not clear to actors on the ground. There are reports of farmers selling off their new lands to plantation companies, although technically the social forestry permit does not give them this type of right.

Second, social forestry permits reduce the risk of expropriation and thus make future returns to the landholder more certain (Goldstein and Udry 2008; Mendelsohn 1994). Land titled with a social forestry permit is less likely to be subject for instance to a future concession claim by a plantation company. Titles are for a period of 35 years. Social forestry permits thus increase the security of investments in the land. On the one hand, this creates additional incentives to restore the land, harvest forests more sustainably, and invest in measures to intensify on existing agricultural land outside the forest. On the other hand, if demand for agricultural products from cleared land is sufficiently elastic, tenure security can also increase the incentive to invest in land clearing. Rakatama and Pandit (Rakatama and Pandit 2020) describe contexts in which social forestry permits have led to expectations regarding increased future land titling, thereby incentivizing communities to claim additional land and attracting additional people towards forest lands. Land-clearing and cultivation has been described as an avenue to claim resource rights in the Indonesian context (The World Bank 2020). This effect would be similar to the "race to property titles" initially described in the context of the Indonesian rubber economy in the 1990s (Angelsen 1995).

Third, social forestry permits may improve community members' access to capital because they can now formally prove that investments they make in their land will be more safe from expropriation and also that they have an additional secure source of income based on forest resources. Community members could increase their investments either in conservation-related activities or in agriculture based on land-clearing with loans enabled by their increased credit-worthiness. An additional mechanism that can increase investments through the access to credit is when land can serve as a collateral (Besley 1995). However, a social forestry permit cannot in practice play this role. Overall, there is no evidence for additional lending activity from private banks. However, the *Funding Analysis Forestry Division* at the Ministry of Environment and Forestry provides loans to permit-holders without background checks (Shahab 2018). This may make investments in activities that are aligned with the social forestry regulations more likely, however it does not in practice rule out investments in land clearing either.

Empirical literature

Reviews on land tenure

Our study examines a large-scale program that grants communities limited rights to forest use in Indonesia. There is a wider literature on the role of land tenure security, including individual rights to agricultural lands, for social and environmental outcomes. A recent review of 117 studies on land tenure security interventions finds that a majority of those studies that looked at forest outcomes found positive effects (Tseng et al. 2020). However, in a recent restrictive meta-analysis that compares interventions targeted at forest conservation land tenure programs (n=4) generate on average worse (heterogeneous) results with the direction of the effect being unclear (Börner et al. 2020).

Prior results on the environmental and social effects of the Indonesian social forestry program

Domain experts note a lack of counterfactual-based impact evaluation of the social forestry program and report heterogeneous results from the field research-based literature (Meijaard et al. 2020; Rakatama and Pandit 2020). Previous counterfactual-based research evaluates the performance of HD permits in Sumatra and Kalimantan between 2012 and 2016, and finds that HD permits appear to have performed well in terms of higher avoided deforestation rates (Santika et al. 2017). It also finds heterogeneous effects on poverty reductions (Santika et al. 2019). Older studies on HKm permits (Djamhuri 2008; Pender et al. 2008) find that HKm permits are associated with more tree planting. There is also early evidence, that social forestry may help increase women's participation in decision-making and the likelihood of girls getting more educational opportunities (Siscawati 2020).

Methods appendix

Calculations: people in and around state forest

We use maps of the Indonesian state forest (*Kawasan Hutan*) and the High Resolution Settlement Layer (Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University 2016) to estimate the number in and at a 1km distance from state forest land. Similar maps and estimates have recently been produced at a lower (30″ × 30″) spatial resolution (Newton et al. 2020; Schleicher et al. 2019).

Note that the map of Indonesian state forest has recently undergone substantial changes. Our estimates are in line with prior estimates in the literature. It has for instance previously been estimated that over a third of villages in Indonesia today are located on forest estate (CIFOR 2019). However, due to the "release" of land from the state forest estate for agriculture, it is likely that our estimate will become increasingly biased upwards.

Stacked difference-in-differences

We build on recent advances in difference-in-differences designs with staggered treatment (Goodman-Bacon 2018) and construct clean control groups for each year cohort of land titles. In standard panel regressions treated units also serve as control units for other cohorts before and after they get treated. This can violate the common trends assumption because of dynamic treatment effects or anticipation. Our research design allows us to restrict the control group to control units that do not get treated themselves. We stack these cohorts into a pooled dataset for our regression analysis.

Econometric specifications

We estimate the following model:

$$\ln \text{Deforestation rate}_{iyc} = \lambda_i + \mu_y + \beta \text{Title}_{ic} + \gamma \text{Post}_{yc} + \delta \text{Title}_{ic} \times \text{Post}_{yc} + \nu \text{Precipitation}_{iy} + \epsilon_{iyc} \quad (4.1)$$

where the parameter of interest is δ, the change in deforestation rate linked to treatment. Units of observation are titled areas and control areas indexed by i. Observations are from a year y and belong to a cohort c. Title is a dummy that indicates whether a study area is treated in the treatment year of a cohort. Post is binary indicator that is 1 if year y is in or after the treatment year of the cohort in which the observation appears and 0 otherwise.

The regressions include a fixed effect in event time γ, that indicates if an observation is before or after the treatment year of a cohort. This fixed effect controls for dynamics before and after a land title roll-out that are common to the whole sample. β captures any systematic differences between the pooled treatment and control group.

Like in a standard two-way fixed effects design, we also use fixed effects at the area (λ_i) and year (μ_y) levels. Thereby we control for time-invariant differences between study areas (for instance topography, climate, institutions) and changes that are common to the whole sample in calendar time, such as macro-shocks or changes in national policy. Our results are robust to the inclusion of fixed effects at the province-year level but standard errors increase (see Robustness section).

We control for precipitation (CHIRPS) (Funk et al. 2015) to rule out that El Niño effects are driving our average effects.

Robustness

In this section we report on the main robustness checks we have conducted. In the text we report on the effects on all types of forests in HD areas. Robustness checks for all types of titles and all types of forests combined are included in the code in the replication package: https://zenodo.org/record/4314768.

Our results are robust to the inclusion of a fixed effect at the province-year level (HD, all forest: CI 95% [-0.02,0.24]), which controls for potential bias from political dynamics at the province level, for instance governors driving both land titling and deforestation outcomes. The results are also robust to removing years between the initial titling step and the final permit in the sample of HD areas, for which we know the year of initiation of the process (all forest: CI 95% [0.08,0.53]). We also confirm our results, when we only keep the areas with the highest level of initial protection (HL, watershed protection forest; in HD areas, all forest: CI 95% [-0.01,0.61]) or for the two highest levels of initial protection (HL, watershed protection forest, and HPT, limited production forest; in HD area, all forest: CI 95% [0.01, 0.47]). Further, our results do not substantially change, when we control for spillovers (SUTVA violations) by (i) excluding areas from the same district from the control group (HD, all forest: CI 95% [0.01,0.26]) or (ii) using only treated units from districts that have not had a treated area before (HD, all forest: CI 95% [-.02,0.32]). In further robustness checks we also restrict our sample to cohorts from 2009 and 2015 (HD, all forest: CI 95% [0.15,0.82]), drop observations from the El Niño year 2015 (HD, all forest: CI 95% [0.05,0.30]), and drop observations after 2016 (HD, all forest: CI 95% [0.13,0.84]) to make our sample period more comparable to previous research (Santika et al. 2017).

4.6 References

Angelsen, Arild (1995). "Shifting Cultivation and "Deforestation": A Study from Indonesia". *World Development* 23.10, pp. 1713–1729. DOI: 10.1016/0305-750X(95)00070-S.

Angelsen, Arild, Pamela Jagger, Ronnie Babigumira, Brian Belcher, Nicholas J. Hogarth, Simone Bauch, Jan Börner, Carsten Smith-Hall, and Sven Wunder (2014). "Environmental Income and Rural Livelihoods: A Global-Comparative Analysis". *World Development*. Forests, Livelihoods, and Conservation 64, S12–S28. DOI: 10.1016/j.worlddev.2014.03.006.

Angelsen, Arild and David Kaimowitz (1999). "Rethinking the Causes of Deforestation: Lessons from Economic Models". *The World Bank Research Observer* 14.1, pp. 73–98. DOI: 10.1093/wbro/14.1.73.

Besley, Timothy (1995). "Property Rights and Investment Incentives: Theory and Evidence from Ghana". *Journal of Political Economy* 103.5, pp. 903–937. DOI: 10.1086/262008.

Börner, Jan, Dario Schulz, Sven Wunder, and Alexander Pfaff (2020). "The Effectiveness of Forest Conservation Policies and Programs". *Annual Review of Resource Economics* 12.1, pp. 45–64. DOI: 10.1146/annurev-resource-110119-025703.

Busch, Jonah, Kalifi Ferretti-Gallon, Jens Engelmann, Max Wright, Kemen G. Austin, Fred Stolle, Svetlana Turubanova, Peter V. Potapov, Belinda Margono, Matthew C. Hansen, and Alessandro Baccini (2015). "Reductions in Emissions from Deforestation from Indonesia's Moratorium on New Oil Palm, Timber, and Logging Concessions". *Proceedings of the National Academy of Sciences* 112.5, pp. 1328–1333. DOI: 10.1073/pnas.1412514112.

Carlson, Kimberly M., Robert Heilmayr, Holly K. Gibbs, Praveen Noojipady, David N. Burns, Douglas C. Morton, Nathalie F. Walker, Gary D. Paoli, and Claire Kremen (2018). "Effect of Oil Palm Sustainability Certification on Deforestation and Fire in Indonesia". *Proceedings of the National Academy of Sciences* 115.1, pp. 121–126. DOI: 10.1073/pnas.1704728114.

Chhatre, Ashwini and Arun Agrawal (2009). "Trade-Offs and Synergies between Carbon Storage and Livelihood Benefits from Forest Commons". *Proceedings of the National Academy of Sciences* 106.42, pp. 17667–17670. DOI: 10.1073/pnas.0905308106.

CIFOR (2019). *Forest tenure reform in Indonesia*. CIFOR. https://www.cifor.org/gcs-tenure/research/research-sites/indonesia/.

Dermawan, Ahmad, Heru Komarudin, and Sian McGrath (2006). "Decentralization in Indonesia's forestry sector: Is it over? What comes next?" Eleventh Biennial Global Conference of The International Association for the Study of Common Property (IASCP). Bali.

Djamhuri, Tri Lestari (2008). "Community participation in a social forestry program in Central Java, Indonesia: the effect of incentive structure and social capital". *Agroforestry Systems* 74.1, pp. 83–96. DOI: 10.1007/s10457-008-9150-5.

Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University (2016). *High Resolution Settlement Layer (HRSL)*. Tech. rep. Source imagery for HRSL © 2016 DigitalGlobe.

Ferraro, Paul J. and Rhita Simorangkir (2020). "Conditional Cash Transfers to Alleviate Poverty Also Reduced Deforestation in Indonesia". *Science Advances* 6.24, eaaz1298.

Fisher, M.R., M. Moeliono, A. Mulyana, E.L. Yuliani, A. Adriadi, Kamaluddin, J Judda, and M.A.K. Sahide (2018). "Assessing the New Social Forestry Project in Indonesia: Recognition, Livelihood and Conservation?" *International Forestry Review* 20.3, pp. 346–361. DOI: 10.1505/146554818824063014.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, and Joel Michaelsen (2015). "The Climate Hazards Infrared Precipitation with Stations—a New Environmental Record for Monitoring Extremes". *Scientific Data* 2.1 (1). DOI: 10.1038/sdata.2015.66.

Goldstein, Markus and Christopher Udry (2008). "The Profits of Power: Land Rights and Agricultural Investment in Ghana". *Journal of Political Economy* 116.6, pp. 981–1022. DOI: 10.1086/595561.

Goodman-Bacon, Andrew (2018). *Difference-in-Differences with Variation in Treatment Timing*. Working Paper 25018. National Bureau of Economic Research. DOI: 10.3386/w25018.

Hajjar, Reem, Johan A. Oldekop, Peter Cronkleton, Peter Newton, Aaron JM Russell, and Wen Zhou (2020). "A Global Analysis of the Social and Environmental Outcomes of Community Forests". *Nature Sustainability*, pp. 1–9.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend (2013). "High-Resolution Global Maps of 21st-Century Forest Cover Change". *Science* 342.6160, pp. 850–853. DOI: 10.1126/science.1244693.

Koch, Nicolas, Erasmus K. H. J. zu Ermgassen, Johanna Wehkamp, Francisco J. B. Oliveira Filho, and Gregor Schwerhoff (2019). "Agricultural Productivity and Forest Conservation: Evidence from the Brazilian Amazon". *American Journal of Agricultural Economics* 101.3, pp. 919–940. DOI: 10.1093/ajae/aay110.

Langston, James D., Rebecca A. Riggs, Yazid Sururi, Terry Sunderland, and Muhammad Munawir (2017). "Estate Crops More Attractive than Community Forests in West Kalimantan, Indonesia". *Land* 6.1 (1), p. 12. DOI: 10.3390/land6010012.

Margono, Belinda Arunarwati, Peter V. Potapov, Svetlana Turubanova, Fred Stolle, and Matthew C. Hansen (2014). "Primary Forest Cover Loss in Indonesia over 2000–2012". *Nature Climate Change* 4.8, pp. 730–735. DOI: 10.1038/nclimate2277.

Meijaard, Erik, Truly Santika, Kerrie A. Wilson, Sugeng Budiharta, Ahmad Kusworo, Elizabeth A. Law, Rachel Friedman, Joseph A. Hutabarat, Tito P. Indrawan, Julie Sherman, Freya A. V. St John, and Matthew J. Struebig (2020). "Toward Improved Impact Evaluation of Community Forest Management in Indonesia". *Conservation Science and Practice* n/a, e2189. DOI: 10.1111/csp2.189.

Mendelsohn, Robert (1994). "Property Rights and Tropical Deforestation". *Oxford economic papers*, pp. 750–756.

Ministry of Environment and Forestry Indonesia (2021a). *Pengembangan Usaha Perhutanan Sosial*. WebGIS. http://sinav.usahahutan.id/index.php/frontend/. Jakarta, Indonesia.

– (2021b). *Peta Indicative Area Perhutanan Sosial*. WebGIS. http://webgis.menlhk.go.id:8080/kemenhut/index.php/id/peta/petapiaps. Jakarta, Indonesia.

Newton, Peter, Andrew T. Kinzer, Daniel C. Miller, Johan A. Oldekop, and Arun Agrawal (2020). "The Number and Spatial Distribution of Forest-Proximate People Globally". *One Earth* 3.3, pp. 363–370. DOI: 10.1016/j.oneear.2020.08.016.

Ostrom, Elinor (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

Pender, John, Suyanto, John Kerr, and Edward Kato (2008). *Impacts of the Hutan kamasyarakatan social forestry program in the Sumberjaya watershed, West Lampung district of Sumatra, Indonesia*. IFPRI Discussion Paper 00769. International Food Policy Research Institute.

Rakatama, Ari and Ram Pandit (2020). "Reviewing Social Forestry Schemes in Indonesia: Opportunities and Challenges". *Forest Policy and Economics* 111, p. 102052. DOI: 10.1016/j.forpol.2019.102052.

Resosudarmo, Ida Aju Pradnja, Luca Tacconi, Sean Sloan, Faridh Almuhayat Uhib Hamdani, Subarudi, Iis Alviya, and Muhammad Zahrul Muttaqin (2019). "Indonesia's Land Reform: Implications for Local Livelihoods and Climate Change". *Forest Policy and Economics* 108, p. 101903. DOI: 10.1016/j.forpol.2019.04.007.

Santika, Truly, Erik Meijaard, Sugeng Budiharta, Elizabeth A. Law, Ahmad Kusworo, Joseph A. Hutabarat, Tito P. Indrawan, Matthew Struebig, Sugeng Raharjo, Imanul Huda, Sulhani, Andini D. Ekaputri, Soni Trison, Madeleine Stigner, and Kerrie A. Wilson (2017). "Community Forest Management in Indonesia: Avoided Deforestation in the Context of Anthropogenic and Climate Complexities". *Global Environmental Change* 46, pp. 60–71. DOI: 10.1016/j.gloenvcha.2017.08.002.

Santika, Truly, Kerrie A. Wilson, Sugeng Budiharta, Elizabeth A. Law, Tun Min Poh, Marc Ancrenaz, Matthew J. Struebig, and Erik Meijaard (2019). "Does Oil Palm Agriculture Help Alleviate Poverty? A Multidimensional Counterfactual Assessment of Oil Palm Development in Indonesia". *World Development* 120, pp. 105–117. DOI: 10.1016/j.worlddev.2019.04.012.

Schleicher, Judith, Julie G. Zaehringer, Constance Fastré, Bhaskar Vira, Piero Visconti, and Chris Sandbrook (2019). "Protecting Half of the Planet Could Directly Affect over One Billion People". *Nature Sustainability* 2.12 (12), pp. 1094–1096. DOI: 10.1038/s41893-019-0423-y.

Shahab, Nabiha (2018). *Taking Stock of Indonesia's Social Forestry Program*. Tech. rep. CIFOR Forest News.

Siscawati, M. (2020). *Gender and Forest Tenure Reform in Indonesia*. Tech. rep. Center for International Forestry Research (CIFOR). DOI: 10.17528/cifor/007572.

Siscawati, Mia, Mani Ram Banjade, Nining Liswanti, Tuti Herawati, Esther Mwangi, Christine Wulandari, Martina Tjoa, and Thomas Silaya (2017). *Overview of forest tenure reforms in Indonesia*. Working Paper 223. Bogor, Indonesia: Center for International Forestry Research (CIFOR), p. 45.

Sunderlin, William D., Arild Angelsen, Brian Belcher, Paul Burgers, Robert Nasi, Levania Santoso, and Sven Wunder (2005). "Livelihoods, Forests, and Conservation in Developing Countries: An Overview". *World Development* 33.9, pp. 1383–1402. DOI: 10.1016/j.worlddev.2004.10.004.

The World Bank (2020). *Strengthening of Social Forestry in Indonesia (P165742)*. Project Appraisal Document PAD3214. Washington, DC.

Tseng, Tzu-Wei Joy, Brian E Robinson, Marc F Bellemare, Ariel BenYishay, Allen Blackman, Timothy Boucher, Malcolm Childress, Margaret B Holland, Timm Kroeger, Benjamin Linkow, et al. (2020). "Influence of land tenure interventions on human well-being and environmental outcomes". *Nature Sustainability* 4, pp. 242–251.

Wijaya, Arief, Nirarta Samadhi, and Juliane Reidinar (2019). *Indonesia Is Reducing Deforestation, but Problem Areas Remain*. Tech. rep. World Resources Institute (WRI).

World Resources Institute, Rainforest Alliance, Proforest, Daemeter, Trase, Earthworm, Auriga, CIFOR, Transitions, Jason Jon Benedict, Robert Heilmayr, and Kimberly M. Carlson (2019). *Universal Mill List*. URL: https://data.globalforestwatch.org/datasets/5c026d553ff049a585b90c3b1d53d4f5_34 (visited on 12/14/2020).

Wunder, Sven, Jan Börner, Driss Ezzine-de-Blas, Sarah Feder, and Stefano Pagiola (2020). "Payments for Environmental Services: Past Performance and Pending Potentials". *Annual Review of Resource Economics* 12.1, pp. 209–234. DOI: 10.1146/annurev-resource-100518-094206.

Wunder, Sven, Jan Börner, Gerald Shively, and Miriam Wyman (2014). "Safety Nets, Gap Filling and Forests: A Global-Comparative Perspective". *World Development*. Forests, Livelihoods, and Conservation 64, S29–S42. DOI: 10.1016/j.worlddev.2014.03.005.

Chapter 5

Price incentives and unmonitored deforestation: Evidence from Indonesian palm oil mills[†,‡]

Abstract

We estimate price elasticities of deforestation across the Indonesian oil palm sector. This is the first study to relate deforestation to observations of the actual prices paid at palm oil mill gates. In the palm oil value chain, mills are pivotal: they have strong influence over deforestation decisions and can be monitored by downstream and public actors. We create a novel, spatially explicit data set of mill-level input and output prices, ownership and export shares from 1998 to 2015. We model the plantation-mill relationships and isolate local, exogenous differences in price incentives at plantation sites. We find that deforestation in industrial, smallholder and illegal plantations is price elastic. This shows that market-based instruments can disincentivize deforestation in the Indonesian palm oil sector, and in particular where it is most difficult to monitor. Our finding that legal deforestation is contrastingly not price elastic suggests that economic opportunities encourage regulation bypassing. This implies that price instruments can help regulations be more effective in this context.

5.1 Introduction

Tropical deforestation due to oil palm plantations is a major source of global biodiversity loss and accounts for 5% of the global green house gas (GHG) emissions since 1986 (Hsiao 2020). Agricultural prices are one of the main drivers of tropical deforestation (Busch and Ferretti-Gallon 2017; Leblois et al. 2017). Regulators need local insights into the responsiveness of deforestation to palm oil prices to design environmental policies or to predict land use change leakages from their jurisdictions to the tropics through commodity markets (Hertel 2018). Yet, for Indonesia, the biggest producing country, it remains unclear, how deforestation actors in the palm oil supply chain react to the price incentives that actually pass-through to them.

Do prices influence deforestation, and which segments of the oil palm sector react more or less strongly to prices? We estimate price elasticities of deforestation across the Indonesian oil palm sector. We build the first spatially explicit data set of prices paid at palm oil mill gates from 1998 to 2015. We find that a 1% increase in crude palm oil (CPO) prices

[†]Valentin Guye and Sebastian Kraus (2021). "Price incentives and unmonitored deforestation: Evidence from Indonesian palm oil mills". In preparation for resubmission at the *Journal of Environmental Economics and Management*.

[‡]We thank Raja Chakir, Sabine Fuss, Jérémie Gignoux, Robert Heilmayr, Nicolas Koch, François Libois and seminar participants at UMR Economie Publique and MCC WG3 for their helpful comments. We are grateful to Jason Benedict and Robert Heilmayr for their support with the Universal Mill List. We thank Claudia Guenther and Hanif Kusuma Wardani for their invaluable research assistance.

over the preceding 4 years increases the average annual conversion of primary forest to oil palms by 1.6%. We call this the medium-run price elasticity. Looking at different segments of the oil palm sector, we find that industrial and smallholder plantations, as well as illegal plantations have a positive price elasticity of deforestation. This constitutes evidence that less institutionalized actors of the sector can be incentivized away from deforestation. Moreover, we find persistent evidence that legal deforestation is not price elastic. This suggests the existence of leakages from legal to illegal deforestation in presence of economic opportunities and weak law enforcement.

Data on the Indonesian palm oil supply chain available to researchers has been limited. Previously, only the location and total capacity of most mills was known and the establishment date for a subset of them. We are the first to geo-localize palm oil mills based on the full Indonesian manufacturing census (IBS).[1] For approximately half of all known Indonesian mills, we observe in particular annual input (palm fruits) and output (crude palm oil) farm gate prices, public, private and foreign ownership shares, as well as CPO export shares. We use data based on satellite imagery to measure deforestation around mills at a high resolution. We detect deforestation as 30 m-pixel events of primary forest loss, conditional on eventual oil palm plantation development. Industrial and smallholder plantations typically differ in scale and landscape. We define illegal deforestation as occurring outside a known concession and inside a state forest zone. Our sample for estimation is a 2002–2014 annual panel of 3×3 km plantation sites in Sumatra and Kalimantan, where most of deforestation due to oil palm plantations occurred during the period.

Our estimation strategy grounds on the fact that palm fruits deteriorate quickly from harvest to processing. This means that each plantation can only reach a limited number of surrounding mills in time. For each plantation site that can reach at least one mill we average the crude palm oil prices of reachable mills. Closer mills get higher weights. These inverse-distance weights model the relative influences of reachable mills on plantations in a way that is consistent with the palm oil sector's heterogeneity: the weights represent either the odds of being integrated plantation-mill systems or transport costs. We average the annual, weighted prices over the four past years to capture medium-run variations we consider most relevant to perennial crop expansion.

The identification of the causal effect of prices on deforestation in our observational setting is subject to several threats. A third unobserved factor may partly explain the observed correlation between prices and deforestation. Typically, sites more suitable for oil palms in lowlands may also be closer to ports and cities and hence experience higher prices. We control away such time-invariant confounders with plantation site fixed effects. Districts are critical jurisdictions in the administration of land in Indonesia. Therefore, local political cycles as well as shocks on the local labor, land and energy markets may confound our estimates. We address this potential source of confounding with district-year fixed-effects. We also control for the number of reachable mills and for their public, private and foreign

[1] In the economics literature this dataset has also been referred to as *Statistik Industri*, see for instance Amiti and Konings (2007).

ownership shares. This helps us to further shelter our estimated effect from biases that could arise from partial equilibrium and political economy effects that may be present even below the district-year level.

A bias could also arise if part of the correlation of interest is actually explained by a reverse effect of deforestation on prices. We leverage the granularity of our data and assume that the size of plantation sites relative to quantities traded on local markets ensures that individual plantations are price taker. Bias due to measurement errors is mitigated because we observe the actual prices paid at mills' gates and model how they are perceived from each plantation site.

Our main results are twofold. First, we find that deforestation in both industrial and smallholder plantations is price elastic. Second, we document that illegal deforestation is price elastic whereas legal deforestation is not. Together, these results have two main implications.

The first implication is that segments of the oil palm sector that are more difficult to regulate – illegal industrial or smallholder plantations – can be incentivized away from deforestation. Tackling land use regulation bypassing with price incentives is critical for conservation policies as new oil palm frontiers – in the island of Papua in particular – seem to largely involve illegal deforestation.[2] This is also relevant to sustainable development policies, because monitoring the encroachment of smallholders on forests is increasingly challenging in Indonesia. To address these challenges, recent fiscal conservation policy proposals have devised a taxation on defaults whereby a commodity tax is uniformly levied at choke points (like palm oil mills), but can be refunded against proof of sustainable production (Heine et al. 2020). Our results indicate that such schemes can work.

Second, our results have implications for regulatory conservation frameworks in similar contexts. Our finding that legal deforestation is inelastic to prices suggests that legal deforestation does not react to medium-run market signals because of long licensing processes. On the other hand, we estimate a substantial price elasticity of illegal deforestation. This indicates the existence of strong incentives to circumvent land use regulations in order to seize economic opportunities of palm expansion. Together, these results imply that more stringent conservatory regulations may extend the licensing process and, in absence of strong monitoring, thus encourage illegal deforestation in the presence of high price incentives. However, this leakage effect can be contained if price incentives are controlled. Hence, our results suggest that in this context a market-based instrument may help conservatory regulations be more effective.

We provide three additional pieces of evidence to better understand how prices affect deforestation. The first one shows that, unsurprisingly, price incentives do drive immediate conversion (within 4 years) of primary forest into oil palms, but not transitional deforestation, in which plantation development occurs 4 years or more after forest clearing.

[2] `https://news.mongabay.com/2018/11/the-secret-deal-to-destroy-paradise/`

Yet, we also find a clear price elasticity of transitional and illegal forest conversion to industrial plantations. This delayed setup of illegal plantations could represent companies that, motivated by palm oil prices, clear the forest, but then face delays in plantation development because of conflicts with communities or legal and bureaucratic proceedings.

Second, we disentangle the effects of short- and medium-run crude palm oil prices and find that short-run (annual) prices alone do not drive deforestation, but strengthen the medium-run price elasticity. This indicates that growers of the perennial and yield lagging oil palm crop look at short-run price signals only to confirm medium-run dynamics.

Third, we disentangle the effects of palm fruit and crude palm oil prices. We document that deforestation in industrial plantations is actually mainly driven by palm fruit prices which vary consequentially to crude palm oil prices. Since the output of vertically integrated plantations (i.e., plantation-mill systems) is crude palm oil while the output of independent plantations is the palm fruit, this finding highlights the role of the latter in price-driven deforestation. Lastly, we find that deforestation in smallholder plantations decreases with palm fruit prices and increases with crude palm oil prices. This suggests that it is the mill owners – usually companies – willing to benefit from higher output/input price ratios, that decide upon the timing and location of smallholder encroachment on forests.

The medium-run crude palm oil price elasticity across the overall Indonesian oil palm sector is 1.6. It is robust in a range of alternative specifications, including the maximum plantation-mill distance, control sets (additionally including lagged deforestation, shares of crude palm oil exported and baseline forest cover trends), regional-year fixed-effects, and standard error clustering levels.

A 1.6 price elasticity of deforestation implies that a 19% tax on crude palm oil can curb deforestation 29% (Indonesia's targeted reduction in GHG emissions under the Paris Agreement) below the 2002–2014 average. We quantify that for the whole country, this represent 39kha of avoided conversion of primary forest to oil palm plantations annually. Under a result-based payment scheme, like the United Nations program REDD+ (Reducing Emissions from Deforestation and Degradation), applying a US\$5/tCO_2 price, this corresponds to US\$124M of yearly revenues. Recycling tax revenues to compensate plantations proving avoided deforestation (as suggested in Heine et al. (2020)) could increase further conservation incentives and outcomes.

Furthermore, the empirical price elasticity estimates we provide can be used to calibrate and assess models that predict land use change and how deforestation reacts for instance to increased biofuel subsidies or trade policies. In this perspective, providing a crop and region specific price response and breaking it down to different economic agents can help refine land use change modelling (Wicke et al. 2012). Deriving our estimation from data on actual prices observed by agents and tackling concerns of endogeneity can also participate to performing more accurate predictions. We emphasize that providing medium-run elasticities is relevant to land use predictions simulating policies that have a lasting effect

on prices[3] (such as taxes, subsidies or tariffs) (Berry 2011).

Related literature This paper principally contributes to the literature shedding light on the economic incentives of land use change.[4] The role of prices in oil palm-related deforestation is a case of particular interest, as indicated by recent efforts in the Global Forest Review to relate time series of deforestation and palm oil prices (Goldman et al. 2020) (see also Gaveau et al. (2021)). Yet, data availability has constrained the identification of causal relationships as well as heterogeneity and mechanism analyses. Thanks to the new spatially explicit microeconomic data we produce and to recent remote-sensing data sets, we are able to advance the literature in the following directions:

We provide the first price elasticity estimates specific to smallholders[5] and illegal oil palm deforestation. These results relate our work to the field of development economics.

We are also the first to explore how short-run prices and palm fruit prices (FFB) affect deforestation and how they interact with medium-run crude palm oil prices in doing so. Methodologically, we are the first to estimate country-level elasticities with actual price observations in the oil palm context. Other studies, using imputed price measures, provide estimates that can be interpreted similarly to some of ours – the elasticity in industrial plantations in particular. The present study advances this literature since it measures actual prices, which allows for improved identification of the price elasticity. In the following, we explain the challenges to identifying a local price elasticity in the previous literature.

Wheeler et al. (2013) were the first to establish a positive correlation between time series of palm oil futures prices and forest loss alerts at a monthly rate. Subsequent studies have advanced the causal price effect identification by adding spatial variation. They proxied local farm prices by interacting international prices with measures of local suitability for palm plantations (Busch et al. 2015; Cisneros et al. 2020; Hsiao 2020). Yet, in these studies the price elasticity of deforestation is not the main parameter, and thus their authors may have naturally focused less on identification concerns about it. Their approach is prone to several sources of bias, that our mill-gate price data helps remove.

First, the suitability-price interaction proxy can be subject to measurement error, including systematic one. For instance, it is possible that more independent (i.e., less vertically integrated) plantations that have longer pass-throughs from international to actual price (Zant et al. 2004) take also systematically different deforestation decisions. Alternatively, potential yields – the measure of suitability – may be a systematically more precise measure

[3]Yet, short-run elasticities also provided in this paper can be useful to dynamic models and/or simulations of punctual market shocks.

[4]This is a large literature and we point in particular to (Busch and Ferretti-Gallon 2017) for a review; (Souza Rodrigues 2019) in the Amazon context; and (Leblois et al. 2017) for a cross-country analysis.

[5]The closest literature on smallholders, based on survey data, estimates a positive correlation between crude palm oil and local land prices (Krishna et al. 2017) and opportunity costs of conservation (Cacho et al. 2014).

of exposure to international prices for particular oil palm actors with specific deforestation patterns (Woittiez et al. 2017). In contrast, our analysis relies on the actual prices paid at mills' gates and to model how they are perceived consistently with heterogeneity in plantation-mill integration. Moreover, this observational level is relevant because mills are pivotal in the palm oil value chain: they are the most influential actors over plantations while they can still be monitored by downstream corporate or public actors. (Purnomo et al. 2018).

Second, the interaction term of suitability and international price creates shocks that conflate price variation with suitability variables including their lagged effects on the levels and growth rates of social and economic variables. Some of these lagged effects can also be due to long-term endogeneous sorting of units between locations because of their different suitability for various economic activities. While suitability measured by climate and soil is largely exogenous to deforestation, its lagged effects are confounding the effect of the individual, contemporaneous price shock. This means that the coefficient on the interaction term could be predominantly driven by suitability and therefore cannot be easily interpreted as a deforestation-price elasticity.

Another source of bias in prior approaches is that individual palm oil districts might not be price-takers. Palm oil production is concentrated in Malaysia and Indonesia and production expansions are typically focused on a small number of areas at a time. As part of the decentralization process Indonesian district governments have held important power in land planning and licensing. Prior research has shown that district splits (Burgess et al. 2012) and political cycles (Cisneros et al. 2020) influence the land supply. Therefore, there might be reverse causality between local deforestation dynamics and international palm oil prices or omitted variable bias related to political economy or local infrastructure. Our mill-level prices provide a level of disaggregation below the district-level, which makes us more confident that in our empirical setup individual plantation owners can be considered price-takers.

Finally, let us remark that the estimates we provide are specific to primary forest conversion to oil palms, and hence exclude deforestation in broader senses (such as not imputable to oil palm plantations or in already degraded forest[6]), which is not always the case in previous comparable research.

The remainder of this paper is organized as follows: Section 2 provides background elements and a theoretical framework both necessary to better understand our empirical model. In Section 3 we present our empirical framework in three parts: the plantation-mill relationship model, the estimation strategy and the identification strategy. Section 4 introduces our new palm oil mill data set and the remote-sensing land use data sets we use; then it provides descriptive statistics for the final sample of plantation sites used in analyses. In section 5 we present and discuss our main estimates, a mechanism analysis, a comparison with previous comparable estimates, and finally scaled-up counterfactuals.

[6]See (Hansen et al. 2014) for a discussion on the use of the Global Forest Change data to study deforestation.

Section 6 concludes. Tables, figures, references and appendices follow in this order.

5.2 Background and theoretical framework

Background: determinants of oil palm expansion and prices in Indonesia

General background Indonesia is the largest producer of crude palm oil (CPO) worldwide since 2006; it produces half of the world total (Byerlee et al. 2016). Exports represent 13% of Indonesian exports (Pacheco et al. 2017) and 75% of domestic production. Because of its numerous applications, from food and cosmetics to biofuels, global demand grew by 7% annually between 1980 and 2013 (Cramb and McCarthy 2016). Domestic demand is expected to increasingly come from the Government of Indonesia's (GoI) biodiesel mandate policy. This agricultural boom has relied on a rapid expansion: the planted area increased by more than 600% between 1990 and 2010 (MoA 2010). Because oil palm is a highly profitable crop in Indonesia, with large suitable areas still available (Pirker et al. 2016), and because installed processing capacities are far from saturated (Pirard et al. 2020), this expansion is not expected to stop.

Drivers of deforestation Not all plantation expansion is associated with deforestation, but a significant part is (Austin et al. 2019). Deforestation has followed different trends across Indonesian islands - Sumatra, Kalimantan and Papua (Austin et al. 2017) - and actors - industrial or smallholder plantations (Lee et al. 2014). The determinants of this deforestation are agro-climatic suitability - elevation, slope, precipitation, soil type - (Byerlee et al. 2016), the proximity to existing plantations (Gunarso et al. 2013; Shevade and Loboda 2019) and to roads (Hughes 2018), the decentralization of authority on land (The Gecko Project[7]; Burgess et al. 2012) and local political cycles (Cisneros et al. 2020). It has been shown that oil palm-related deforestation follows global prices of crude palm oil (Busch et al. 2015; Gaveau et al. 2018; Wheeler et al. 2013; Zikri 2009).

Drivers of local prices Local prices are partially determined by international prices, which, beside global demand for palm oil, depend on large scale meteorological events like El Niño (Rahman et al. 2013), the Indonesian export tax (Rifin 2014) and substitutable commodity prices (Sanders et al. 2014; Santeramo and Searle 2019). Yet, prices may vary locally, due to several factors. Local prices of palm fruits, commonly called fresh fruit bunches (FFB), are supposed to be collectively determined by provincial governments, firms and farmers. However, it has been shown that they actually result from each mill's discretionary decisions based on its monopsonic market power, the quality of FFB purchased, and each mill's CPO sales (Maryadi et al. 2004; Masliani et al. 2014).

[7] https://thegeckoproject.org/

"Farm gate" CPO prices, the unitary values paid to each mill for its CPO, are driven by international prices, domestic prices, the export tax, and each mill's CPO quality and bargaining power with downstream refineries or traders. The latter depends, in particular, on the share and the nature of off-take agreements each mill has with downstream buyers. As shown in Pirard et al. (2020), there is little integration between mills and refineries, and the ties between operating companies are mostly hidden from the public as of now.

Overview of oil palm plantation-mill systems

This sub-section gives an overview of the diversity of production systems around oil palm plantations and mills in Indonesia since the late 1990s. We provide a general description of the setup of crude palm oil production, and then present industrial and smallholder plantations, making further typical distinctions within each group.

The plantation-mill proximity constraint Oil palm trees produce fruits (often referred to as fresh fruit bunches, or FFB) that can be harvested around 10 times a year, for around twenty years. Once harvested, these fruits damage quickly because they rot fast and bruise easily during transport. The fruits are brought by trucks and/or by river boats to factories, called mills, that process them into crude palm oil. The quality and quantity of oil derived from a tonne of fruit increase with the quality of the fruits and thus decrease with the distance from the trees to the mill (Byerlee et al. 2016). This constraint leads to spatial proximity between mills and plantations.

Mills need a minimal fruit supply basis to operate. At usual mill capacity and plantation yield, this implies a minimum plantation size of ca. 3000 hectares to be developed alongside any new mill opening (Paoli et al. 2013). As the first palm fruits can be harvested three years after planting, plantation development starts before the mill starts operating. Because of the inverse relationship between distance and fruit quality, these plantations are developed in the immediate vicinity of the new mill. We refer to this as plantation expansion at the extensive margin. Once a mill is established, the plantation-mill proximity constraint is relaxed to the broader surrounding area and more plantations are developed - either by the same company or by other actors. We call this expansion at the intensive margin.

Industrial plantations Industrial plantations are large grid-shaped landscapes ranging from a hundred hectares to hundreds of thousands of hectares (Austin et al. 2017; Gaveau et al. 2016). They represent the majority of planted area and production in Indonesia. They are developed by companies or public governments but here, for simplicity, we refer to a 'firm' as the agent deciding on the timing and location to develop an industrial plantation. Some industrial plantations are integrated with mills and sometimes also further downstream with refineries and exporters, but, in the light of the best knowledge

of the field, this integration seems limited (Pirard et al. 2020). Hence, industrial plantations are heterogeneous in how they sell their fruits, from internal transactions to partial off-take agreements, to selling on the local spot market of mills they can reach in time.

Smallholders The term 'smallholder' lacks a common definition, but is often used in contrast with some or all of the characteristics of industrial plantations presented above. In this study, we refer to smallholder plantations to broadly designate small and medium sized plantations, developed in mosaic landscapes (i.e., alongside other land uses, such as subsistence crops, in particular). These smallholder plantations may belong to individuals, households, cooperatives, or companies. They are heterogeneous in their sizes, land ownership, management, relationship with companies' mills and industrial plantations.[8] The main distinction usually made is between independent smallholders and those that are part of a scheme with a larger structure. This distinction is relevant to our study because it affects how decisions on timing and location of smallholder plantations are affected by prices.

Broadly, supported smallholder plantations (also called *plasma*) are developed jointly with a firm aiming at developing industrial plantations (called *inti* or *nucleus*). The timing and location of this expansion results from this firm's decisions (Paoli et al. 2013) and the fruits have to be sold to the firm's mill. Independent smallholder plantations are developed outside of negotiations, without direct influence from a firm. Their fruits are sold on the local spot markets, through off-take agreements with middlemen, or directly "at the gate" of the mill.[9]

Overall, beyond the aforementioned grouping characteristics, Indonesian oil palm plantations differ in sizes, ownership, management and marketing. They are developed at the extensive or at the intensive margin, and they replace diverse previous land uses, in different degrees of legality.

Theoretical framework

This subsection sets some term definitions for the present paper and then explains how we assume deforestation to react to prices.

Definitions Throughout this paper, we use the term *plantations* to designate micro-economic agents deciding where and when to clear forest for the purpose of planting

[8]See Baudoin et al. (2017) and Cramb and McCarthy (2016) for more insights into the diversity of smallholders.

[9]See Euler et al. (2016) and Jelsma et al. (2017) for further insights into independent smallholders.

oil palms.[10] It is important to note that, in this study, we refer to plantations as agents prospecting to plant oil palms and not as a realized land use.

We use the term *deforestation* to refer to land use change from forests to oil palms. More conceptually, it is any forest clearance that is motivated by the intention to grow oil palms. Hence, our use of this term here excludes any other forest loss phenomenon.

The term *price signal* refers to the information that plantations use to anticipate future cash-flows associated with deforestation.

Reduced form assumptions First, we assume that deforestation results from decisions taken by plantations. The typical decision rule is the comparison of the expected discounted present utilities (or profits) from alternative inter-temporal scenarios defined by the timing and the amount of deforestation. Hence, the counterfactual scenario includes any other use of forested land (conservation or alternative land use). A scenario with zero deforestation corresponds to two possible situations: either the plantation decides to expand only outside forests, or it decides to not expand at all (i.e., to not enter the market as a plantation). We do not distinguish between these two situations in our analysis.

Plantations are assumed to form expectations on the basis of privately observed informational elements (Stavins 1999). Beside price signals, we expect plantations' decisions to be influenced by information related to investment costs (e.g., of land acquisition and conversion), operating costs (e.g., of labor, energy, and fertilizers), institutional costs (either fixed or marginal, positive or negative, formal or not) and attainable yields. Plantations also take into account the expected relative costs and benefits of the alternative land uses. Because oil palm is a perennial crop, the land use change decision is highly committing. Plantations formulate medium-run expectations. This implies that their discount rates and abilities to make expectations are also parameters of the decision rule.

Conceptually, this may all be summarized in the reduced form relationship between deforestation and price signals:

$$Deforestation = f(PS, U) \tag{5.1}$$

where PS is the true price signal that enters the representative plantation's decision rule on deforestation, conditional on other structural error terms U, including the information elements mentioned above.

We do not attempt to formally model how complete information sets determine deforestation decisions. Therefore, instead of specifying a deterministic form for the function f, we

[10] We purposely do not refer to "landholders", "landowners", "growers" or "farmers", in order to abstract as much as possible from notions of ownership, legality, or management. This seeks generality over the diversity of actors that may be involved in the decision process towards a plantation development.

approximate it - as explained in the next section.

5.3 Empirical framework

This section first provides information on our empirical unit of observation. Then, we introduce our efforts to empirically model the diversity of plantation-mill systems described in Section 5.2. We then present our regression equation and strategy to estimate the reduced form introduced in Section 5.2. Finally, we explain our causal identification strategy.

Empirical definition of the theoretical micro-economic agent In our theoretical framework, the typical micro-economic agent is an oil palm plantation. Empirically, we do not observe the actual boundaries between plantations. Thus, we approximate the theoretical individual plantations with square land parcels of an equal size. Each year, deforestation in each of these grid cells is assumed to result from decisions taken by an homogeneous, profit maximizing plantation agent. In this paper, we refer to plantations as the micro-economic unit of interest, although what we actually observe are grid cells. We choose the typical size of grid cells to be 3×3 km (900ha). This is the outcome of a trade-off: it is small enough to assume that the average plantation area in a grid cell is price taker on its palm fruit market, but large enough to limit spatial auto-correlation and keep computation times reasonable.[11]

Measuring price signals: an empirical model of the plantation-mill relationship

For every plantation, every year, there is a true, privately known price signal. For mills, we know the annual "farm gate" prices for crude palm oil (CPO). For plantations, we approximate the true price signal with a mix of these prices at the mills the plantation can reach in time, before the fruit spoils. What constitutes this mix has implications for how the different segments of the oil palm sector contribute to our estimation. The next four paragraphs explain how.

The set of reachable palm oil mills For each plantation, we determine a set of reachable mills for each year. Mills are considered reachable if they are within a circular area around the plantation, determined by a catchment radius parameter. We assume that freshly harvested palm fruits can potentially be brought to any mills within this area without deteriorating too much. Mills beyond the catchment radius are not reachable and thus are assumed to have no influence on the plantation's decision to deforest.

[11] This is also the size of grid cells in Busch et al. (2015).

In this study, our preferred catchment radius is 30 km in Sumatra and 50 km in Kalimantan. Choosing the value for this parameter results from a trade-off. On the one hand, a too short catchment radius implies observing too few of the plantations experiencing deforestation and biasing our observations towards areas near palm oil mills. On the other hand, a large catchment radius implies spuriously relating plantations to more mills that, despite being reachable, are actually unrelated. This would, in turn, make our price elasticity estimate less precise. This trade-off justifies that we assume a different catchment radius for Sumatra and Kalimantan. First, in Sumatra, typically most deforestation occurs within 30 km of mills, while in Kalimantan a significant share occurs farther away (see Table 5.1). Second, the higher mill concentration in Sumatra reduces the likelihood that a plantation will be influenced by prices from mills located farther than 30 km away.

The existing literature helps us get a sense of magnitudes for catchment radii of palm oil mills. According to Harris et al. (2013), only 15.3% of oil palm plantations are farther than 30 km from a mill.[12] The Center for International Forestry Research (CIFOR), in its online atlas[13] applies a 10 km buffer around mills. In Peninsular Malaysia, a region comparable to Sumatra, Shevade and Loboda (2019) report almost no deforestation due to oil palms beyond 40 km to a mill.

Table 5.1: Deforestation accumulated over 2002-2014, in kha

	Sample	30 km from sample mill	50 km from sample mill	Total
Sumatra	220.29	564.55	702.02	801.40
Kalimantan	150.32	321.92	565.81	1,015.62
Both	370.62	886.47	1,267.83	1,817.02

Notes. This table shows measures of accumulated deforestation from 2002 to 2014 in different groups of Indonesian plantation sites. Deforestation is counted as primary forest loss eventually (by 2015) replaced with oil palm plantations (either industrial or smallholders). The sample of plantation sites is the one we actually use in estimations. Sample mills are the 587 palm oil processing plants from the Indonesian manufacturing census that we have geo-localized.

Mill influence intensities A plantation can reach several mills, but not all mills are equally influential. We do not directly observe how prices paid at every reachable mill enter the price signal that is observed by each plantation. Therefore, we model these intensities using straight line distances between each plantation and its annual set of reachable mills. More precisely, we model the price signal[14] as the standardized invert-distance weighted average of "farm gate" prices at reachable mills.

As depicted in Section 5.2, both industrial and smallholder plantations may be vertically integrated to different extents, from full integration with one mill, to having partial off-take agreements, to selling on the spot market only (full independence). We do not observe the

[12]The study is based on Gunarso et al. (2013) for plantation data and Global Forest Watch for palm oil mill data, for Indonesia, Malaysia, and Papua New Guinea. 44.5% of oil palm plantations are within 10 km of a mill, and 8.1% are farther than 50 km.

[13]`https://atlas.cifor.org/borneo/#en`

[14]This is explained here for the price signal, our explanatory variable of interest, but the same method is applied to all mill-level covariates.

degree of integration of each plantation. Yet, the standardized inverse-distance weights enable modeling of the relative influence from reachable mills in a way that is consistent over degrees of integration. To see this, consider two main types of plantations: those selling exclusively to (and hence getting a price signal from) one mill, and those selling at least some of their outputs on the local spot market, i.e., to any reachable mill (and getting a composite price signal). Plantations in the former category are typically close to the mill they sell to. Thus, the standardized inverse-distance weights approximate the odds to be integrated with each reachable mill. For plantations in the second category, the standardized inverse-distance weights approximate the expected transport costs to every reachable mill (including fuel costs and fruit quality decline). Prices at mills relatively farther away are less influential, because reaching them from the plantation site is more costly.

Prices of palm fruits and prices of crude palm oil We know the annual average prices offered at mills' gates for fresh fruit bunches (FFB), as well as annual average prices received at mills' gates for crude palm oil. Prices of FFB and prices of CPO are assumed to affect deforestation decisions differently, depending on the degree of plantation independence. More independent plantations tend to look more at FFB prices, while more integrated plantations tend to look more at the prices of CPO (the output of the plantation-mill integrated system). Besides, mills are assumed potential price makers on the FFB market, but not on the CPO market. Therefore, at the mill level, CPO prices may affect FFB prices and thus indirectly influence independent plantations. On the other hand, FFB prices do not affect CPO prices and thus do not influence integrated plantations. In order to include integrated plantations in the analyses, we mainly focus on CPO price signals. Moreover, FFB price signals are likely more endogenous to deforestation than CPO's, making the price elasticity identification with FFB price variations less robust. Finally, potential price instruments are more conceivable at the more downstream level of CPO market and thus CPO price elasticities are more relevant to policy implications.

Measures of price signals over time Because oil palm is a perennial crop, developing a plantation is a committing decision. Thus, short-term (annual in our case) changes in prices alone are unlikely to motivate deforestation. To allow deforestation to react to slower price variations, our main measure is the 4-year average of annual price signals.

Besides, we lag all variables from the Indonesian manufacturing census by one year. This merely aims at correcting a measurement lag. We do this because remotely sensed annual deforestation does not necessarily represent the actual state at the end of the year, while IBS variables should, a priori, reflect census respondents' observations for the whole year. Because this does not have conceptual implications for our empirical strategy, we do not annotate these lags or refer to them further.

Estimation strategy

Here, we present the empirical model estimated in this study. We also discuss the assumptions we rely on for inference.

Regression equation In this study, we estimate the following model, which is an empirical approximation of equation (5.1).

$$Deforestation_{idt} = exp(\alpha ln(Price_{idt}) + \beta X_{idt} + \lambda_{id} + \gamma_{dt} + e_{idt}) \tag{5.2}$$

From year 2002 to 2014 ($t = 1, ..., 14$), we observe $Deforestation_{idt}$, which is the sum of pixel-level deforestation events[15] in plantation i in district d in year t. Hence, $Deforestation_{idt}$ is a count of non-negative integers that may be null for a significant proportion of observations and we assume it follows a quasi-Poisson distribution (Wooldridge 1999) (see Appendix 5.7 for more details).

$Price_{idt}$ is a measure (detailed below) of the price signal observed by plantation i in district d in year t. α is the price elasticity of deforestation. X_{idt} is a vector of other observed determinants of deforestation that vary both locally and annually. In our main specification, X_{idt} comprises measures of the share of domestic private capital and of the share of foreign capital (we exclude the share of public capital to avoid perfect collinearity). It also includes a measure of local market development (see Section 5.3 for more detail on the role of control variables). The unobserved determinants of deforestation can be decomposed as a sum of heterogeneity sources that can be either fixed attributes of plantations (local fixed effects, λid), or annual shocks common to a whole district (district-year fixed effects, γ_{dt}), or idiosyncratic error terms, e_{idt}. We elaborate on observed and unobserved heterogeneity in the next subsection.

Clustering We do not assume that annual records of price signals are independent and identically distributed. Rather, we allow arbitrary correlations within clusters of observations. Abadie et al. (2017) explain that clusters should be set at the level of randomization. In our case, as we do not use experimental data, identifying the proper clustering level is not straightforward. The minimal effort is to cluster at the level of the unit fixed effect (i.e., letting individual prices be correlated over time). Yet, this is not sufficient because price signals are also spatially dependent. This is due to the fact that price signals in neighboring plantations only differ in their sets of reachable mills and distances to these. To address this spatial auto-correlation, we have to cluster at a broader

[15]Each grid cell is composed of 30m x 30m pixels that were either forested or not in 2000 ($t = 0$). A deforestation pixel event, as explained in Section 5.4, is when forest loss is observed in a pixel where oil palms are subsequently observed.

level than the cross-sectional unit. Like Dell and Olken (2020), in a study of Indonesian sugarcane mills, we cluster at the level of the sub-district.

Identification strategy

The causal interpretation of the observed correlation between prices and deforestation - identification of the price elasticity - is threatened by reverse causality, omitted variable bias and measurement error. Reverse causality can arise, for instance, if deforestation increases the palm oil supply, or expectations about it, pushing prices downwards. This is even more likely in the presence of spatial auto-correlation in deforestation. A third variable could also drive both prices and deforestation, biasing the causal interpretation of the observed correlation. This could, for instance, be due to local politicians opening up land for plantations and creating new infrastructure. Finally, measurement error, random or systematic, may also lead to spurious causal conclusions. It is possible, for instance, that the international price is a more precise measure of the true price incentive for plantations integrated in large companies, which also present systematically different deforestation patterns.

We mitigate these threats to identification by using high-resolution, local variations of deforestation, prices and control variables. We control away the confounding co-variations of prices and deforestation that are constant between plantation sites, but also those between districts each year. Indonesian districts have important powers in palm oil licensing and are, therefore, the critical geographic level at which both omitted variable and reverse causality biases can arise. We condition our identifying comparisons to the ownership of reachable mills. This helps us further shelter our price elasticity estimates from some political economy effects that may be present even below the district level. Finally, we rule out potentially confounding competition and partial equilibrium effects below the district level with a proxy for local oil palm market development. This is especially relevant as palm oil may be the backbone of the local economy in some Indonesian regions, with disparities even within districts.

Intuitively, the remaining identifying variation occurs each year within each district, as cross-plantation comparisons of how much their yearly deforestation and price signals deviate from usual. We assume that differences are due to independent combinations of relative distances to mills and mill-level price shocks from downstream (due, for instance, to changes in contracts with, or costs of transport to, exporters or refiners). We argue that these shocks, conditional on the controls and fixed effects, are indeed independent from the joint spatial distributions of mills and plantations. To summarize, we assume that, conditional on constant features, annual district shocks, within district market development and reachable mills' ownership, the price signals that plantations get to observe are as-good-as-randomly assigned. In the following, we discuss how we address the three types of identification threats in more detail.

Reverse causality In our setting, reverse causality could arise and bias our estimates, if deforestation affected prices. It is conceivable, indeed, that deforestation leads to increased production and hence affects prices. As explained above, our price signal variable is a 4-year average. In addition, there is a 3-year time lapse between planting and first harvests and oil palm trees are not always planted immediately after forest clearing. This rather long time lag between our measure of price signal and the potential reverse effect of deforestation on prices is a first argument against the presence of reverse causality. However, as demonstrated in Bellemare et al. (2017), such an argument relies on the assumption that there are no dynamics in the confounders. In our case, deforestation may, indeed, be correlated over time. Past deforestation may cause more production and lower prices now, while also affecting (positively or negatively) present deforestation. Our argument against this potential channel of reverse causality is the size of one plantation grid cell, our cross-sectional unit of observation (see Section 5.3). In this study, a grid cell has an area of 900 hectares - approximately 0.3% of the area around a mill with a catchment radius of 30 km. This is even less if one thinks about larger markets than only one mill, like developed palm fruit markets with several mills, or the CPO market. Hence, the core identifying assumption here is that annual deforestation by an individual plantation is generally marginal enough not to affect mills' prices. In economic terms, we assume that plantations are price-takers.

Spatial auto-correlation Spatial auto-correlation can lead to a particular case of reverse causality. Given the 3×3 km spatial resolution in our data, our deforestation observations may be spatially auto-correlated.[16] We may also expect that deforestation in neighboring plantation sites 4 years later aggregate and gather enough market power to impact local prices. To rule this threat out, in the robustness analysis in Appendix 5.7 we check that controlling for the 4-year lagged average deforestation in neighboring plantation sites does not change our main estimate.

Omitted variable bias We use mill features available from the manufacturing census and a spatial measure of the local market development as controls, as well as plantation and district-year fixed effects, to limit omitted variable bias. First, we control for reachable mill average ownership.[17] We include in X_{idt} the share of domestic private capital and the share of foreign capital (we exclude the share of public capital). Ownership controls capture systematic correlation between deforestation and prices following capital shifts (either investments in a new mill or purchase) across public, private and foreign origin. We believe this control to be important as, for instance, local government plantations may have different deforestation motivations than foreign plantations, while also having different marketing conditions. Second, we control for the number of known mills reachable

[16]Evidence of significant positive spatial auto-correlation from Costa Rica (for forest loss Robalino and Pfaff (2012)) and Malaysia (for expansion of oil palm plantations Shevade and Loboda (2019)) justify such concerns.

[17]With the same standardized inverse-distance weights as described in Section 5.3.

from each plantation annually. This captures systematic differences between frontier and mature markets. We refer to this control here as *local market development*. The higher the mill density, the more developed the local markets for plantation and mill inputs like land, labor and palm fruits in particular, and the more developed the local infrastructure. Local market development is likely to have an effect on both local prices and deforestation. Moreover, the mill density is also a proxy for plantations' market power because, in high mill-density areas, independent plantations have a higher bargaining power and FFB prices are higher (Maryadi et al. 2004; Masliani et al. 2014).

Plantation fixed-effects remove any time-invariant heterogeneity. Notably, this prevents agro-climatic and geographic heterogeneity from confounding our estimates. Such spatial heterogeneity may determine potential yields and hence deforestation, and could also be correlated with determinants of transport costs (distance to refineries or exporters) and institutional costs (distance to cities can proxy the intensity of monitoring by law or civil society). District-year fixed-effects capture economic and political shocks down to the district level. Districts are powerful jurisdictions in the administration of land in Indonesia. The control over land can unlock substantial revenues from natural resources. Therefore, political cycles at the district level can explain much of deforestation.[18] At the same time, many other determinants of prices can vary annually at this level, through the local input (e.g., labor, land and energy) and output markets (including FFB, which prices are supposedly determined at the provincial, i.e., upper, level).

Measurement error. We believe that our data and estimation strategy enable us to get the most accurate measure of the true price incentives privately observed by oil palm plantations in Indonesia to date. However, some measurement error remains. Here are its main sources: First, we observe the annual mean unitary values and not the prices that mills publicly disclose. Second, we can only model the price signal that reaches individual plantations (cf. Section 5.3). Third, our sample of geo-localized IBS mills does not cover the whole population. Therefore, in areas with mills both from and not from our sample, our measure of the price signal is incomplete. We do not suspect any of these to be prone to systematic measurement error. In particular, Table 5.8 shows that there is no systematic difference between the IBS mills we have geo-localized and the others.

5.4 Data and descriptive statistics

In this section, we present the data we use to measure the components of Equation (5.2) above. The first subsection documents our original micro-economic dataset of geo-localized palm oil mills. The methodology to measure price signals and transform the mill data into the final sample of plantations is not described here but in Section 5.3. The

[18]Indeed, district splits (Burgess et al. 2012) and competition for district head election (Cisneros et al. 2020) have been shown to be determinants of deforestation.

second subsection presents the land use data, along with the methodology to measure deforestation. The third subsection describes the final sample.

Micro-economic data: an original merge of the Indonesian manufacturing census and the Universal Mill List

We matched two existing data sets to produce an original spatially explicit microeconomic data set of palm oil mills in Indonesia from 1998 to 2015.

Indonesian manufacturing census (IBS) The Indonesian manufacturing census (IBS)[19] is issued by the Indonesian office of statistics (BPS). It reports annual establishment-level data for all manufacturing facilities employing at least 20 employees. We identified palm oil mills with 9-digit commodity codes[20] from 1998 to 2015. The variables available in the manufacturing census and used in our analysis are geographic variables;[21] mill-level input and output quantities and values at the 9-digit commodity level; mill-level ownership shares across four categories (national public, regional public, domestic private and foreign private); and product-level export shares.

We identified 930 plants as palm oil mills, based on the criteria that they sourced FFB at least once or sold CPO or PKO at least once, and that they are not located in Java or in Bali.

We computed the mean unitary values of inputs (FFB) and outputs (CPO) by dividing values by quantities. We expressed all monetary values used in the analysis in 2010 USD. In the paper, we call these values "farm gate" prices of FFB and CPO. Input-output variables, as well as village identifiers are usually not provided to researchers with the manufacturing census. However, these variables are essential in building the spatially explicit price data used in this paper.

Universal Mill List (UML) In the latest version we use, the Universal Mill List features 1140 Indonesian palm oil mills, with their names and coordinates (UML 2018). We merge the UML with a newer data set of palm oil mills (Heilmayr et al. 2020), containing information on parent companies and establishment dates, but we further refer to the whole data set as the UML.

Matching the manufacturing census and the UML We matched the palm oil mills from these two data sets to make the manufacturing census' economic data spatially explicit.

[19]The data has also been referred to as *Statistik Industri* in the literature.

[20]KKI codes used are 151410102, 151410103 for crude palm oil and crude palm kernel oil respectively, and 011340101 or 011340501 for fresh fruit bunches.

[21]The data we obtained from BPS provided the district (*kabupaten*) information over the 1998-2015 period. However, the sub-district (kecamatan) and the village (desa) information were provided over 1998-2010 only.

The matching strategy leverages a third document: the manufacturing directories. This is a list of manufacturing establishments, with their names, 5-digit industry codes, main commodity names, addresses (often incomplete), and number of workers. Although they are edited annually, we could find them only for years 2003, 2006, 2009-2015. Since the number of workers in the directories is sourced from the manufacturing census,[22] we used this variable together with district (and village when available) information to match mills from the manufacturing census with manufacturing directories' names. These names were then used to match the manufacturing census mills with UML coordinates. All conflicts were resolved after a case-by-case investigation. Finally, we match 466 mills from the manufacturing census with a UML palm oil mill (and four more which never reported CPO or PKO output, nor FFB input, or are located in Java).

There are 464 palm oil mills from the manufacturing census that could not be matched with the UML by the method explained above. Out of these, we approximate the geo-localization of the 121 additional mills for which village information is reported in the manufacturing census. To do so, we use the centroids of the polygons of the most recent valid village identifier. Because, in Indonesia, since 2000, there is a trend to village splits rather than to village mergers, the most recent information also tends to be the most spatially accurate.[23]

The final, spatially explicit mill sample is constituted of 587 palm oil mills. 466 of them are matched with a mill referenced in the UML and hence have exact coordinates, and 121 are not matched with the UML but are approximately geo-localized at their village centroids.

Table 5.8 shows descriptive statistics of Indonesian palm oil mills, along with evidence that the subset of these mills used in the present analysis is not significantly different from the overall sample of palm oil mills in the manufacturing census. It is important to recall that in this paper, mills are not the cross-sectional units of observation. The mill-level price and ownership data are used to compute the variables at the level of grid cells, as explained in Section 5.3. See Section 5.4 for a description of the grid cell sample and variables used in estimations.

Land use change from forest to oil palm plantations

In this section, we explain how we construct our measures of land use change from forest to oil palm plantation (referred to as 'deforestation' here).[24]

[22] Although with many lags, leads, and inconsistencies between the two.

[23] Due to administrative village splits, plants do not necessarily report their correct village names or codes every year. This can be particularly misleading because codes for "parent" villages may be re-used in the next iteration but for different villages than their "child" villages. Therefore, we deemed that the village information a plant reported in a given year was valid if the corresponding "parent" village (in 2000) matched with the mode of all annual village information reported by the plant (also expressed in "parent" village).

[24] All rasters used in this study are aligned with the resolution of forest loss maps from Hansen et al. (2013) and all spatial data are projected with a Cylindrical Equal Area projection centered on Indonesia (longitude = 115, latitude = 0).

Forest loss We use maps from the Global Forest Change (GFC) dataset (Hansen et al. 2013). They cover the whole of Indonesia with a resolution of 1 arc-second per pixel (i.e., approximately 30 meters per pixel in our near-equator region of interest)[25] annually from 2001 to 2018. A forest loss event is defined at the pixel level, as the year when complete removal of tree (with a minimum height of 5m) canopy cover is observed where such cover was still present in 2000. A minimum canopy cover threshold defines what is counted as forest in 2000 at the pixel level. However, the GFC dataset does not enable us to distinguish between 2000 tree canopy cover (and hence loss) in primary forest, secondary forest, or tree plantations.

Primary forest extent in 2000 The map we use to measure primary forest extent in 2000 comes from Margono et al. (2014). It covers the whole country, with the same resolution as the GFC data set. Primary forest in 2000 is a subset of the 2000 tree canopy cover from the GFC data set, with canopy cover of at least 30%. It is defined as "mature natural forest cover that has not been completely cleared in recent history and consisted of a contiguous block of 5ha or more" (Margono et al. 2014). Two primary forest types are distinguished: intact and degraded. The former, following Potapov et al. (2008), shows no sign of alteration by humans, while the second has been subjected to human disturbances, such as selective logging. They correspond to the Indonesian Ministry of Forestry's primary and secondary forest cover types (Margono et al. 2014). In this study, we regroup them.

Oil palm plantations In this study, we use two different maps, from Austin et al. (2017) and Petersen et al. (2016). These maps have been produced by visual interpretation of Landsat imagery. They both recognize areas with signs of future cultivation as plantations. The former product, from Austin et al. (2017), includes only large-scale oil palm plantations and covers the regions of Sumatra, Kalimantan, and Papua for the years 1995, 2000, 2005, 2010 and 2015, with a 250m pixel resolution. The latter product, from Petersen et al. (2016), includes and distinguishes between large plantations of more than 100ha, mid-size plantations and small-size plantations. It is a snapshot of the whole of Indonesia, computed with images from 2013 and 2014. Mid and small-size plantations are mosaic landscapes. Mid-size plantation mosaic landscapes are at least 100 hectares wide, have oil palm patches between 10 and 100 hectares, comprising at least 50% of the landscape. Small-size plantation mosaic landscapes have oil palm patches smaller than 10 hectares, again comprising at least 50% of the landscape.

In our main analysis, we use the maps from Austin et al. (2017) to study industrial plantations, and we pool small and mid-sized plantation maps from Petersen et al. (2016) to study smallholder plantations. Where these map sources overlap, we characterize plantations as industrial, as remote sensing for this landscape is less error-prone.

[25] 27.8 x 27.6 meters with our projection.

Measuring deforestation We combine these data sets to compute annual maps of deforestation, which we measure in different ways.

First, we consider two alternative forest definitions at the pixel level, and hence two different forest extents in baseline year 2000: tree canopy cover of at least 30 percent outside of 2000 industrial oil palm plantations (as observed by Austin et al. (2017)),[26] and tree canopy cover of at least 30 percent within all (i.e., intact or degraded) primary forest. The latter is the closest to the official forest definition by the Government of Indonesia (Austin et al. 2017; MoF 2008) and, therefore, the one used for our main analyses.

Then, annual forest loss pixel events observed within the two 2000 baseline forest extents are deemed deforestation events if they later fall within an oil palm plantation. This means that we count a deforestation pixel-event the year the forest is cleared, and not the year the palm trees are planted or when they become productive. Therefore, our observation is close to the moment when the deforestation decision is actually taken, and irrespective of provisional land uses. Such provisional land uses between forest clearance and oil palm planting, however, seem rare (Gaveau et al. 2018).[27] Moreover, note that our approach does not count forest degradation as deforestation, because the tree loss pixel-event is counted only once, the year a near-zero canopy closure is observed (Hansen et al. 2013).

For industrial plantations, we further distinguish between immediate and transitional deforestation. We use the time lapse between the forest loss event and the year when a plantation is observed for the first time in data from Austin et al. (2017). Deforestation is deemed immediate if the time lapse is between 0 and 4 years. It is deemed transitional if the time lapse is between 5 and 14 years. For smallholder plantations (data from Petersen et al. (2016)), we only observe one cross-section for the year 2014 and, hence, we cannot differentiate immediate from transitional deforestation.

Legal and illegal deforestation Observing illegal deforestation in Indonesia is challenging because the line between legality and illegality is blurred by weak institutions (especially in rural areas or outer islands where oil palm has developed) and because data are scarce and often contradictory. In this study, we deem deforestation illegal if it occurs outside a known concession and inside a permanent forest zone designation.[28] The map of concessions is provided by the Indonesia Ministry of Forestry (MoF), Greenpeace, and the World Resource Institute (Greenpeace 2011). However, it is only a screenshot, so it does not specify the date the concessions were issued. Furthermore, this map is known to

[26]This ensures that canopy closure removals within already existing plantations (i.e., palm replacements) are not counted as deforestation. This approach is the best we can do in the absence of other tree plantation maps for 2000, but it sill has some pitfalls. For instance, if an area was covered with another plantation type (like timber) in 2000, cleared and converted to an oil palm plantation before 2015, it would be mistakenly counted as deforestation.

[27]In Borneo, Gaveau et al. (2018) found that 92% of the forest cleared for oil palm plantations was planted with oil palms the same year it was cleared.

[28]Any of KSA, KPA, KSAL, HP, HPT or HL.

be incomplete.[29]

Land designation data are provided by the MoF (2019). They are also a screenshot and hence do not contain information on changes over time (like forest release, for instance).

Estimation sample and descriptive statistics

Estimation sample Our sample is an annual unbalanced panel of 3×3 km grid cells[30] in Sumatra and Kalimantan[31] from 2002 to 2014.[32] Sumatra and Kalimantan are the two main Indonesian regions where oil palm expansion occurred during our study period (Austin et al. 2017).

We further restrict the sample in several dimensions. First, we include only observations of grid cells from years when at least one geo-localized mill from the manufacturing census is reachable. Second, we restrict our analysis to grid cell annual records that have a positive forest extent at the start of the year. The aim is to sample the maximum number of observations where prices can have an influence on deforestation choices. This also makes our Poisson estimation less prone to zero-inflation. Third, we remove observations as soon as they are included in an RSPO certified concession (Carlson et al. 2018). Indeed, we expect the effect of prices on deforestation to be systematically different there from other areas.[33] Fourth, we remove annual records as soon as one of the variables in Equation 5.2 has a missing value. In our case, this has a particular influence on the final sample, because the likelihood that a price signal value is missing decreases with the number of reachable mills. Thus, removing missing values implies that we tend to sample fewer grid cells in remote areas. Another particular implication of removing missing values in our case is that we do not sample records of grid cells in the first 4 years after the first reachable mill is established (as our main, medium-run, price signal measure runs over 4 years). Table 5.2 shows that this does not make a significant difference for the price signals, but does for deforestation and the number of reachable mills. This is not surprising, since inclusion in the sample is a function of the number of reachable mills. We argue that this necessary sampling step does not risk introducing a selection bias, as we control in our regressions precisely for the criterion behind it: the number of reachable mills.

[29] `https://www.arcgis.com/sharing/rest/content/items/f82b539b9b2f495e853670ddc3f0ce68/info/metadata/metadata.xml?format=default&output=html`

[30] Precisely, 27.8x27.6m pixels aggregate to 3002.4x3008.4m grid cells.

[31] We do not include observations from other Indonesian islands, where data is too scarce. In Papua, we have very few observations and in other islands data on oil palm plantation extents are lacking.

[32] Although we have data on year 2015 for industrial plantations, we do not include these observations in order to observe them in the same time period as smallholder plantations. We start observing 4-year average price signals in 2002.

[33] This applies to very few grid cells of our sample because few certifications were issued in the first years of the RSPO, from 2009 to 2014.

Table 5.2: Estimation sample – descriptive statistics

	Without missing values			With missing values			t test	KS test
	# grid cells = 4747 # grid cell-year = 31650			# grid cells = 8355 # grid cell-year = 91499				
	mean	std.dev.	median [min; max]	mean	std.dev.	median [min; max]	p-value	p-value
Deforestation (ha)	11.71	43.72	0.08 [0; 847.5]	12.35	46.74	0 [0; 903.1]	0.027	0.000
Price signal ($/tCPO)	681.4	91.27	686.8 [349.8; 926.4]	681.1	91.04	685.3 [349.8; 926.4]	0.660	0.688
Public ownership (%)	14.71	23.71	0 [0; 100]	12.83	26.99	0 [0; 100]	0.000	0.000
Domestic private ownership (%)	69.98	28.39	75.81 [0; 100]	69.66	35.12	83.57 [0; 100]	0.101	0.000
Foreign ownership (%)	15.3	21.38	5.11 [0; 100]	17.51	28.28	0 [0; 100]	0.000	0.000
# reachable mills	9.45	5.51	8 [1; 37]	7.06	4.92	6 [1; 37]	0.000	0.000

Notes. This table shows descriptive statistics of the variables used in our main regression, for the sample of plantation sites (3×3 km grid cells) actually used in estimations (without missing values), and the same sample but without removing observations with missing values. The two right-most columns show p-values of Welch two-sided t-tests, where the null hypothesis is that the true difference in means between the two groups is null, and the groups' variances are not assumed to be equal; and p-values of Kolmogorov-Smirnov tests where the null hypothesis is that the variables in the two groups are drawn from the same continuous distribution. Price signal and ownership variables at the plantation level are inverse-distance weighted averages of these variables at reachable mills.

Descriptive statistics Table 5.2 provides descriptive statistics from the final sample used in the main estimations of Equation 5.2. We recall briefly the definitions of the variables. Deforestation is a count of pixel-level events of primary forest loss eventually (by 2015) replaced by oil palms. In Table 5.2, it is converted to hectares for better readability. Deforestation is positively skewed, with many zero values, and hence the Poisson distribution is relevant to model it. Price signal is the plantation-level inverse-distance-to-mill weighted average of CPO prices at reachable mills, averaged over the 4 previous years. It is expressed in 2010 USD per ton CPO. It is less dispersed than the mill-gate prices in Table 5.8, since it is an average of these. Public, domestic private and foreign ownership shares are the plantation-level inverse-distance-to-mill weighted averages of the ownership shares of the reachable mills. They are expressed in percentage points. The domestic private ownership is the most prevalent (70% on average), and the public and foreign ownership shares are equivalent (15% each) The number of reachable mills is the annual count of known palm oil mills (as from the UML) within a 30 km (50 km in Kalimantan) catchment radius from a plantation. It ranges from 1 to 37, and half of the observations can reach more than 8 mills.

5.5 Results

This section first introduces and discusses our main results: the price elasticities of deforestation in different segments of the Indonesian palm oil sector. Second, we investigate how price elasticities vary with price dynamics and vertical integration. Third, we compare our estimates to the existing literature. Finally, we discuss the external validity of our results and calculate scaled-up counterfactual effects.

All results are derived from Equation 5.2, estimated as explained in Section 5.3, with data presented in Section 5.4. Recall that in all regressions, unless noted, the outcome,

deforestation, is measured as the count of pixel-events of primary forest loss eventually replaced by an oil palm plantation. The treatment, the price signal perceived by a plantation, is measured as the 4-past-year average of annual inverse-distance weighted averages of CPO prices at the gates of reachable mills. In Appendix 5.7, we explain in more detail how we derive partial effects from regression coefficients. Under the assumptions in Section 5.3, we interpret these partial effects of price signals on deforestation causally and refer to them as price elasticities.

The price elasticity of deforestation in Indonesian oil palm plantations

Table 5.3 shows our estimates of the price elasticity of deforestation for different kinds of oil palm plantations. In Table 5.9, we also display the estimated partial effects of control variables on deforestation. The right-most column in Table 5.3 shows that, pooling together all kinds of plantations, we find a 1.6 medium-run price elasticity of deforestation. More precisely, in places where 4-past-year average prices annually deviate from usual (i.e., from their own local mean over time) by 1%, average deforestation is 1.6% higher. In more intuitive words: a 1% increase in price signals makes plantations increase deforestation by 1.6%. Hence, we find that, overall, deforestation due to oil palms does react to price signals in Indonesia. In Appendix 5.7, we conduct a robustness analysis, summarized in the specification chart presented in Figure 5.1. The next paragraphs, and Tables 5.3 and 5.10, document which subgroups of the Indonesian plantation sector contribute to making this estimate lower and/or less precise, and which do not. Indeed, the magnitudes and precision of price elasticity estimates are heterogeneous over the different segments of deforestation.[34]

Industrial and smallholder plantations Breaking down the estimation into plantation types,[35] we find that to a 1% increase in price signals, industrial and smallholder plantations react with a 2.1% and a 1.5% increase in average deforestation, respectively.

This positive price elasticity of deforestation in industrial plantations indicates that corporate actors of the oil palm sector engage in large-scale deforestation where prices are higher than usual. This suggests that medium-run price signals (over 4 years here) do influence large long-term investments, typically over more than a decade. In the next subsection, we disentangle annual price variations to provide more insights into the dynamics of price signals.

The positive price elasticity of deforestation we estimate in smallholder plantations

[34] In Table 5.11, we also present effects of interactions between the price signal and ownership or local market development covariates. It appears that the price elasticity of deforestation does not substantially depend on these covariates.

[35] As detailed in Section 5.4, the distinction between industrial and smallholder plantations is based on the landscape and size differences between plantations mapped by Austin et al. (2017) and the mid and small-sized plantations mapped by Petersen et al. (2016).

Table 5.3: Price elasticities of deforestation across Indonesian oil palm plantations

	Industrial plantations			Smallholder plantations			All plantations		
	Legal	Illegal	All	Legal	Illegal	All	Legal	Illegal	All
Price elasticity									
Estimate	0.51	5.18	2.11	0.14	2.09	1.53	0.2	3.03	**1.64**
CI[95%]	[-2.02; 3.04]	[1.58; 8.78]	[0.11; 4.11]	[-2.22; 2.5]	[0.71; 3.47]	[0.25; 2.81]	[-1.56; 1.95]	[1.16; 4.9]	**[0.3; 2.98]**
Observations	13081	4951	25249	2971	3412	8611	15139	7848	**31650**
Clusters	218	176	326	79	83	152	236	204	**362**

Notes. This table shows our main estimates of the price elasticity of deforestation. They are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The price signal is measured as the 4-year average of annual inverse-distance weighted averages of crude palm oil prices at the gates of reachable mills. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate industrial from smallholder plantations based on scale and landscape criteria (Austin et al. 2017; Petersen et al. 2016). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

indicates that smaller plantations, organized in mosaic landscapes with other land uses, encroach on forests where prices are higher than usual. This responsiveness to crude palm oil prices suggests that it is actually mill owners - most usually companies - that decide upon the timing and location of smallholder plantation expansion. In the next subsection, we differentiate the effect of palm fruit prices to provide more insights into this direction.

Deforestation in industrial plantations seems more price elastic than in smallholder plantations. The difference is even greater in the case of illegal deforestation, where the point estimate for industrial plantations is more than twice as large as for smallholders. Yet, as documented in Table 5.12, differences between industrial and smallholder plantations are not significant at any conventional level of statistical confidence.

Legal and illegal deforestation We further break down the estimation according to the legal status of deforestation.[36] We find close to zero effects of price signals on legal deforestation, irrespective of the plantation type. On the other hand, illegal deforestation appears to be price elastic in every plantation type. Overall, the price elasticity of illegal deforestation is 3. Industrial and smallholder plantations react to a 1% increase in price signals by illegally deforesting 5.2% and 2.1% more respectively.

The positive price elasticity we estimate for illegal deforestation indicates that economic opportunities encourage plantations to circumvent land use regulations. On the other hand, we estimate that legal deforestation is not price elastic. This may come from a lack of statistical power to detect a true positive price elasticity, or to legal deforestation being truly inelastic to prices. Given the magnitude of the estimate (0.2 across plantation types) and the number of observations and clusters (subdistricts), we believe that it is rather truly

[36] In this paper, we define illegal deforestation as deforestation occurring outside a known concession and inside a permanent forest zone designation - cf. Section 5.4.

inelastic to prices. This is most likely the consequence of the long processes necessary to acquire a plantation license (involving, for example, measuring environmental suitability and community consultation; see Paoli et al. (2013) for more detail on the licensing process). If obtaining the legal green lights to clear the forest and plant palm trees takes several years, and 3 to 4 additional years must then be waited before young trees bear first fruits, it is not surprising that medium-run price signals do not influence legal deforestation. Plantations probably rely on more stable signals than those that we capture in this study to formulate long-term expectations about the profitability of engaging today in legal deforestation.

Table 5.12 documents that the differences in the price elasticities of legal and illegal deforestation are statistically significant (except for smallholders). The estimated price elasticity of illegal deforestation is of larger magnitude, and has wider confidence intervals, than for all deforestation (legal, illegal and unknown combined). This is true for any plantation type (industrial, smallholder, or both). This is especially pronounced for industrial plantations, where the price elasticity point estimate is more than twice as large for illegal deforestation. Altogether, these findings about legal and illegal deforestation indicate that, across plantation types, positive price elasticity is driven by illegal deforestation.

Immediate and transitional deforestation As explained in more detail in Section 5.4, we observe both the moments of forest loss and of planting and, for industrial plantations only, can calculate time lags between the two. We consider deforestation to be transitional if more than 4 years elapse between forest loss and plantation development. Table 5.4 shows our estimates of the price elasticity of immediate and transitional deforestation, again distinguishing legal, illegal, and overall deforestation. Overall, the price elasticity of immediate deforestation (2.7) is larger but not more precisely estimated than for immediate and transitional deforestation taken together (2.1). For transitional deforestation, it is lower (1.9) and even less precisely estimated. However, we note that the price elasticity of transitional deforestation is substantial in illegal deforestation, where it is estimated at 6.9. It is more precisely estimated than the also substantial price elasticity of immediate, illegal deforestation (6.6). Whether immediate or transitional, legal deforestation has a low and imprecise price elasticity estimate.

That immediate deforestation is more sensitive to price signals than transitional deforestation is not surprising from our theoretical point of view. It is expected that higher price signals motivate plantation agents to clear forest and grow oil palms as soon as possible to realize higher profits. In other words, it is not expected that oil palm price signals cause forest clearances that are not intended for immediate oil palm development. In this respect, the large and precise estimate for transitional illegal deforestation is surprising. It suggests that a significant number of industrial plantations have observed price signal incentives to develop oil palms and consequently cleared forest illegally, but then refrained from immediate development. This constitutes a piece of evidence

Table 5.4: Price elasticities of immediate and transitional deforestation in Indonesian industrial plantations

	Immediate conversion			Transitional conversion		
	Legal	Illegal	All	Legal	Illegal	All
Price elasticity						
Estimate	1.23	6.61	2.7	-0.34	6.92	1.95
95% CI	[-1.74; 4.2]	[1.59; 11.63]	[0.22; 5.18]	[-3.77; 3.1]	[3.37; 10.48]	[-1.01; 4.9]
Observations	11308	3959	21629	5945	2185	11704
Clusters	193	156	296	190	125	269

Notes. This table shows our estimates of the price elasticity of deforestation in industrial oil palm plantations. They are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The price signal is measured as the 4-year average of annual inverse-distance weighted averages of crude palm oil prices at the gates of reachable mills. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate immediate from transitional deforestation based on the time lapse between forest loss and plantation development (the cut-off point is 4 years). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

that long-term land use change dynamics may be initiated by more medium-run price incentives. We do not observe, and hence do not investigate further, these transitional dynamics here. One can only hypothesize about the possible mechanisms behind them. Price elastic transitional illegal deforestation may presumably be due to rapidly evolving incentives, or to our measure of price signal capturing long-run (>4 years) expectations. It may also be attributed to companies being incentivized to clear the forest in order to grab land outside oil palm concessions, but then facing delays in plantation development because of conflicts with local communities or legal proceedings.

Spatial heterogeneity Finally, we estimate the price elasticities of deforestation for Sumatra and Kalimantan separately. Table 5.10 shows that, in Kalimantan, there are fewer clusters (subdistricts) and observations than in Sumatra, and thus estimates are less precise. It is also possible that, in Kalimantan, we managed to geo-localize a lower share of the universe of palm oil mills, and thus suffer from more noise in the price signal variable, yielding downward biased estimates. Yet, it is also possible that, during our study period, deforestation in Kalimantan was driven by different dynamics than in Sumatra, and that prices were, indeed, less influential (with a relatively larger role played by political economy factors, for instance).

Vertical integration and signal dynamics

Here, we investigate how the medium-run crude palm oil price signal affects deforestation. We disentangle the price elasticity of deforestation in two dimensions: vertical integration and the time length of price signals. To do so, we use, in turn, two new variables: the medium-run palm fruit price signal and the short-run crude palm oil price signal. Each of them is arguably a post-treatment variable in the sense that they do not affect, but are affected by, the treatment (the medium-run crude palm oil price signal). For the palm fruit price signal, this hinges on the assumption that mills have market power on their input (palm fruit) market, but not on their output (crude palm oil) market. The post-treatment status of the short-run price signal relies on the temporal causal argument that past prices affect current prices, but the reverse is false. Therefore, each of them can have an indirect effect on deforestation, whereby the medium-run crude palm oil price signal affects the post-treatment variable, which then affects deforestation. These post-treatment variables can also have a moderation effect on deforestation, whereby they affect the treatment effect. We estimate the partial effects of the post-treatment variables unconditional and conditional on the treatment. When conditional on the treatment, the partial effects of the post-treatment and of the treatment variables exclude the indirect effect. In any case, the partial effects include the moderated and the unmoderated effects (see Appendix 5.7 for more detail). We report the moderation effects as partial effects of terms of interactions between the post-treatment variable and the treatment.

The causal interpretation of all these partial effects relies on the same identification strategy as presented in Section 5.3: plantation and district-year fixed effects plus ownership and local market development controls. In particular, it relies on the assumption that controls and fixed-effects rule out post-treatment confounders that would affect both the post-treatment variable and deforestation. This assumption may be stronger in the case of palm fruit price signals than short-run price signals. Under these assumptions, the conditional partial effects of the post-treatment and treatment variables can be interpreted as net of the indirect effect. The partial effect of the treatment variables unconditional on the post-treatment variables presented in the previous section can be interpreted as total effects. Thus, the difference with conditional partial effects presented here document indirect effects.

Palm fruit and crude palm oil price signals In this study, our main measure of price signals uses crude palm oil prices (see Section 5.3). Palm tree fruits, commonly called fresh fruit bunches (FFB), are sold by independent plantations to mills. The effect of palm fruit price signals on deforestation may thus document the price elasticity of less vertically integrated plantations. Table 5.5 shows our estimates of palm fruit and crude palm oil price elasticities, along with the partial effects of their interactions on deforestation. Table 5.5 also displays the partial effects of palm fruit price signals unconditional on the effect of crude palm oil prices. All models are based on the same specifications as the main

one, from Equation 5.2. Because the spatial distribution of palm fruit price shocks may be more endogenous to deforestation decisions than that of crude palm oil prices, the identification assumptions are probably stronger in this exercise. Hence, estimates from Table 5.5 should more cautiously be seen as descriptive rather than causal.

Palm fruit price signals seem to influence deforestation, but in opposite directions in industrial and smallholder plantations. In industrial plantations, a palm fruit price increase of 1% causes an increase in average deforestation of 1.8%. On the other hand, in smallholder plantations, it causes a decrease in average deforestation of 2%. Across plantation types, these effects cancel out, but the more numerous industrial plantations weigh more and the elasticity is positive. This pattern is similar whether conditional or not on crude palm oil prices. For any plantation type, crude palm oil price elasticity estimates are close to zero once the effect of palm fruit price signals on deforestation is taken into account. The interaction partial effect on deforestation is positive. This means that the effect of crude palm oil price signals on deforestation increases with palm fruit price signals (and vice-versa).

Table 5.5: Palm fruit and crude palm oil price elasticities of deforestation across Indonesian oil palm plantations

	Industrial plantations		Smallholder plantations		All plantations	
FFB price signal						
Estimate	1.82	2.95	-1.97	-1.43	1.28	2.07
95% CI	[0.04; 3.59]	[0.28; 5.63]	[-3.24; -0.69]	[-2.71; -0.15]	[-0.05; 2.6]	[0.35; 3.8]
CPO price signal						
Estimate		0.86		1.44		0.93
95% CI		[-2.12; 3.83]		[-0.46; 3.35]		[-0.88; 2.74]
Interaction						
Estimate		0.14		0.04		0.09
95% CI		[0.01; 0.27]		[-0.05; 0.12]		[0.01; 0.17]
Observations	22903	17918	8250	7087	29070	23185
Clusters	325	279	143	135	357	309

Notes. This table shows our estimates of the palm fruit and crude palm oil price elasticity of deforestation. They are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The price signal is measured as the 4-year average of annual inverse-distance weighted averages of either palm fruit or crude palm oil prices at the gates of reachable mills. The last block of rows shows estimates of the partial effects of the interaction of both, evaluated at the sample mean. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate industrial from smallholder plantations based on scale and landscape criteria (Austin et al. 2017; Petersen et al. 2016). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

In industrial plantations, the bulk of the effect of crude palm oil price signals on deforestation (as estimated in our main analysis, see Table 5.3) is actually attributable to the mechanism of local crude palm oil prices influencing local palm fruit prices which, in turn, affects deforestation decisions. This suggests that deforestation in industrial

plantations occurs mainly in independent plantations - presumably as a result of the low vertical integration, even in the sector's downstream part (Pirard et al. 2020). The positive interaction effect indicates that palm fruit price elasticity is even larger where crude palm oil price signals are high. This suggests that higher crude palm oil prices reinforce expectations about high palm fruit prices and hence motivate deforestation.

In smallholder plantations, our results indicate that deforestation increases in times and places of low palm fruit prices but high crude palm oil prices. This suggests that it is the companies owning the mills, wishing to benefit from higher output/input price ratios, that decide upon the timing and location of smallholder plantations. In the case of plasma smallholders, this is a known fact, but in the case of independent smallholders it is less clear. Since independent smallholders have driven smallholding plantation development since the 2000s, they should also drive our results. Hence, these results further suggest that the expansion of independent smallholders onto forests is driven by mill-level decisions.

Short-run and medium-run price signals The short-run price signal is the inverse-distance weighted average of prices at reachable mills' gates in the same year as the outcome, deforestation. The medium-run price signal, the main measure of the treatment variable in this study, averages short-run price signals over the four past years (see Section 5.3). Table 5.6 shows our estimates of short- and medium-run price elasticities, along with the partial effects of their interactions on deforestation. Table 5.6 also displays the partial effects of short-run price signals alone - i.e., not conditional on the effect of medium-run price signals. All models are based on the same specifications as the main one, from Equation 5.2.

Short-run price signals alone do not explain deforestation. However, once medium-run price signals are included in the model, the partial effects in the short-run increase substantially (except for smallholders). At constant short-run price signals, the effects of the medium-run price signals are lower than without conditioning to short-run price signals (as in Table 5.3).[37] The interaction partial effect on deforestation is positive. This means that the effect of medium-run price signals on deforestation increases with short-run price signals (and vice-versa).

These results may reflect the fact that more recent developments in prices weigh more on expectations and hence on deforestation decisions than older prices. Moreover, it seems that short-run prices influence deforestation only when longer variations are also accounted for. This is at least partly due to the positive moderating effect of short-run price signals on medium-run ones. Together, these results suggest that, given the perennial and yield-lagging features of the oil palm crop, plantations look at short-run price signals only to confirm medium-run dynamics. It is also notable that this pattern comes from industrial plantations, and that it is reversed in smallholder plantations. Indeed, among

[37]This follows mechanically since the first annual price signal is included in the medium-run measure.

Table 5.6: Short-run and medium-run price elasticities of deforestation across Indonesian oil palm plantations

	Industrial plantations		Smallholder plantations		All plantations	
Short-run price signal						
Estimate	0.32	1.11	0.19	0.463	0.3	0.826
95% CI	[-0.14; 0.78]	[0.55; 1.67]	[-0.22; 0.59]	[-0.163; 1.09]	[-0.07; 0.67]	[0.401; 1.252]
Medium-run price signal						
Estimate		0.96		1.23		0.888
95% CI		[-0.521; 2.441]		[0.268; 2.192]		[-0.132; 1.907]
Interaction						
Estimate		0.03		0.042		0.029
95% CI		[-0.005; 0.065]		[0.003; 0.08]		[0.003; 0.05]
Observations	56153	25249	15805	8611	67765	31650
Clusters	454	326	207	152	505	362

Notes. This table shows our estimates of the short- and medium-run price elasticity of deforestation. They are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The short-run price signal is measured as the inverse-distance weighted average of crude palm oil prices at the gates of reachable mills. The medium-run price signal is the 4-year average of short-run price signals. The last block of rows shows estimates of the partial effects of the interaction of both, evaluated at the sample mean. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate industrial from smallholder plantations based on scale and landscape criteria (Austin et al. 2017; Petersen et al. 2016). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

smallholders, the total price signal effect seems to be driven by medium-run variations. One hypothetical explanation for this difference is that, in times of short-run price spikes, companies prioritize deforestation for industrial plantations, and then allocate forest land to smallholder plantation development.

Comparison with existing estimates

Here, we attempt to compare our findings with the closest estimates in the literature. Yet, we remark that none of the studies discussed here have provided a price elasticity of deforestation as their main estimate. Therefore, they may naturally have focused less on identification concerns about this parameter. The first (in time) study we can compare our estimates to, is Wheeler et al. (2013). They estimate a log-log regression of deforestation on a time series of palm oil futures prices and other economic variables. We can compare our estimated price elasticity to their model coefficient of 0.816. Using our spatial variation, we hence find a price elasticity twice as large as theirs. We shall note that this difference may also come from differences in the measure of deforestation between our two studies.

Comparing with Busch et al. (2015) requires more assumptions, because this study provides an estimate of the effect of agricultural revenue - and not price - on deforestation. They

find that an additional $100 (in 2005 USD) is associated with a 1.02-1.18% increase in deforestation. Converting to 2010 USD, assuming an average yield of 3.5 ton CPO per hectare (Khatiwada et al. 2018) and an average price of $680 / ton CPO over the period (based on our own data), we convert their estimates into a 0.13-0.15 price elasticity.[38] This is much lower than our estimated 1.29 price elasticity of deforestation in all industrial plantations (not in primary forest only), which is the most comparable setting to theirs. One should note that the agricultural revenue in Busch et al. (2015) is computed at each land parcel for the most potentially lucrative crop, which is oil palm 69% of the time. Beside this point, one possible explanation of our finding a much larger price elasticity is that our estimation benefits from reduced random measurement error, and hence less bias towards zero.

In Cisneros et al. (2020) the effect of price exposure (calculated as the interaction of international prices and suitability) on deforestation is expressed for one standard deviation. Thus, in order to compare our analyses to theirs, we compute our partial effects for one standard deviation in our data (remaining after fixed-effect variations are absorbed). In their study, a one standard deviation higher palm oil price exposure results in an 8% increase in deforestation. This is exactly equivalent to the effect of one standard deviation in our setting (corresponding to our main 1.6 price elasticity estimate). However, for the two studies to be more aligned, we compare our price elasticity in industrial plantations (10.2% increase in deforestation for a one standard deviation increase in price signals) to their estimated effect of price exposure on deforestation in new industrial oil palm plantations by 2015 (3% and imprecise). Hence, here too, our research setting seems to capture a larger effect of prices on deforestation in the Indonesian oil palm sector. Our findings are also quite divergent in the exercise of comparing immediate and transitional dynamics: while they find respectively a precise 31.3% and -16.5% price exposure effect, we find significant and non-significant 12.9% and 8.5% effects, respectively. Here, the difference from our results may be explained by Cisneros et al. capturing mechanisms at the district level while we use more local variations only.

Scaled-up counterfactuals

In this subsection, we attempt to give a sense of the magnitudes that are implied by our estimated 1.6 micro-level price elasticity of deforestation. First, we discuss the external validity of our results. In light of this, we then describe how we scale up average partial effects. Finally, we present and discuss scaled effects of counterfactual price changes.

External validity Given the specific organisation of the palm oil sector in Indonesia, our results cannot automatically be extrapolated to other crops or countries. Even within

[38] We convert the additional $100 to a $100 * \$100/(0.518 * 3.5 * 680) \approx 8.110924$ percentage change in CPO prices (where 0.518 is approximately the deflator we use). We then scale the associated percentage change in deforestation - either 1.02 or 1.18% - by this relative price change.

Indonesia, given the differences between Sumatra and Kalimantan observed in this study, one should be cautious in extrapolating our results to specific regions like the new deforestation frontier in Papua. However, as the regions in our analysis include most existing Indonesian oil palm plantations and deforestation, we are confident in claiming external validity with respect to the country as a whole. Extending our conclusions in time should also be done with caution, since our study does not cover recent developments in oil palm-related policies such as the biofuel mandates (Kharina et al. 2016) or the *No Deforestation, No Peat, No Exploitation* commitments from the private sector (Pirard et al. 2015). We believe that, although our sample is restricted to plantations within 30 km (50 km in Kalimantan) from at least one mill (to avoid introducing too much noise into our sample), the results can be extrapolated to plantations located even further away. This is supported by our finding that price elasticity is not contingent on our measure of remoteness - the number of reachable mills (Table 5.11). Finally, we note that our estimates mainly capture effects on deforestation at the intensive margin, i.e., occurring after at least one mill opened (see Section 5.2). It is possible that deforestation at the extensive margin is differently sensitive to prices.

Scaling factor To scale up our estimated average price effects to the whole Indonesian country, we count the number of individual plantation sites (grid cells) where deforestation is possible in Sumatra and Kalimantan. Hence, we first count grid cells that are within 82km of at least one known (as from the UML) palm oil mill. This follows Heilmayr et al. (2020), who analyzed from RSPO audit reports that 99% of mills' supply bases were within this straight line distance. Because, in this area, many plantation sites are actually unlikely to experience deforestation (either because there is no forest or because of unsuitability to oil palms), we excluded those that never experienced any deforestation from 2002 to 2014 (as we did to our analysis sample). Note that we count in the scaling area the plantation sites where deforestation occurred before the first mill opened in the catchment radius - i.e., at the extensive margin. Finally, we aggregate our results over 11396 3×3 km plantation sites in Sumatra and Kalimantan. We assume that this population of plantation sites has the same average deforestation as predicted in our sample. Under this assumption, we multiply by the scaling factor to estimate a baseline total deforestation of 134,230 ha.

Counterfactual effects Table 5.7 shows the aggregated annual effects of different counterfactual CPO price changes on deforestation in Indonesia. For a 1% increase in prices, the first column shows the relative change in average deforestation (the price elasticity), the scaled effect on deforestation, and the corresponding potential revenue from a CO_2 payment. The effect is scaled based on the aggregation factor presented above. We estimate corresponding carbon pricing revenues from a potential result-based payment for reducing emissions from deforestation. We apply an average of 638 tCO_2ha^{-1} emissions due to deforestation (Guillaume et al. 2018).[39] CO_2 revenues are based on the \$5/tCO_2

[39]We apply the 44/12 C to CO_2 conversion factor to their 174 $Mg\ C\ ha^{-1}$ lost in conversion of Sumatra rainforests into oil palm monocultures.

agreed price Norway paid to Indonesia for its recently avoided deforestation.[40]

Table 5.7: Counterfactual annual effects of different CPO price changes on deforestation in Indonesia

	+1 std. dev.	-1%	-19%
Relative change (%)	8.31	-1.63	-29
Total change (ha)	11083	-2170	-38699
Potential CO_2 revenues (M$)	-35.4	6.9	123.5

Notes. This table shows scaled-up effects of counterfactual changes in crude palm oil (CPO) price signals. To compute total change effects, we apply relative changes to average predicted deforestation from our main econometric model, with a scaling factor of 11396, equal to the number of 3×3 km grid cells in Sumatra and Kalimantan within 82 km to any known palm oil mill where deforestation occurred at least once between 2002 and 2014. Potential CO_2 revenues correspond to result-based payments paid at a price of $5 per tCO_2 avoided, assuming average emissions of 174tC per hectare deforested.

Hence, given a 1.6 price elasticity of deforestation, we estimate that average variations (+5%)[41] in CPO price signals incentivize Indonesian oil palm plantations to clear 11kha of primary forest annually. In the presence of a result-based payment scheme, this represents a yearly opportunity cost of M$35. To curb annual deforestation 29%[42] below the 2002-2014 average with price incentives alone, price signals for individual plantations should be lowered by 19%. This would save 39kha of primary forest annually, corresponding to revenues from a potential result-based payment scheme of M$124.

5.6 Conclusion

In this study, we estimate different price elasticities of primary forest conversion to oil palm plantations in Indonesia. We find that medium-run CPO price signals have an overall positive effect on deforestation in the Indonesian oil palm sector. The price elasticity is 1.6. Industrial, smallholder and illegal plantations are responsive to prices. On the other hand, price signals have no effects on legal deforestation.

To conclude, we discuss some limitations the reader should be aware of, we present the policy relevance of our results, and propose further research avenues.

Study limitations Our estimates of the price elasticities of smallholders and illegal deforestation are, to the best of our knowledge, the first in the literature on oil palms. Yet, they necessarily rely on observational data that are still scarce and incomplete. This prevents us from ruling out some confounding threats. Notably, the concession data we use to identify legal and illegal deforestation are known not to be exhaustive (see Section

[40] `https://www.regjeringen.no/en/aktuelt/noreg-betaler-530-millionar-for-redusert-avskoging-i-indonesia/id2722135/`

[41] We compute standard deviations in our price signal regressor variable, in the estimating sample, after removing variations at the levels of fixed-effects (Mummolo and Peterson 2018).

[42] Aligning annual deforestation reduction to Indonesian Paris Agreement targets, i.e., 29% GHG emission (including LUC) below business as usual by 2030 (GoI 2016).

5.4). The land zoning data are time-invariant and thus do not inform us about land releases. For these two reasons, we may identify too much illegal deforestation. This imprecision may bias our results if it is correlated locally with prices signals and deforestation. For instance, a district jurisdiction could release forest estate land to oil palm production land in some areas, impacting local palm oil prices there, as well as deforestation. This systematic measurement error would bias the overall estimate.

We also highlight that the external validity of our study may be limited by the exclusion of the extensive margin in our analysis. At the extensive margin, deforestation occurs far from existing mills, and thus local price signals do not exist. We would expect that such deforestation is less price elastic, because it depends more on other elements that determine the mill establishment, like the procurement of a planting license, infrastructure development, or capital availability.

Policy relevance The main implication of our results is that less institutionalized segments of the oil palm sector that are more difficult to regulate - smallholders and illegal plantations - can be incentivized away from deforestation. Smallholders represent an increasing share of palm oil production in Indonesia, while their expansion is harder to monitor than industrial plantations'. A market-based conservation scheme could hence be effective in reducing deforestation inside, but also outside, large-scale, more easily monitored, industrial plantations. Tackling the bypassing of land use regulation is also critical, as new oil palm frontiers - in the island of Papua, in particular - seem to largely involve illegal deforestation.

Furthermore, our finding that legal deforestation is inelastic to prices suggests that legal deforestation does not react to medium-run market signals because of long licensing processes. On the other hand, we estimate a substantial price elasticity of illegal deforestation. This indicates the existence of strong incentives to circumvent land use regulations in order to seize economic opportunities for palm expansion. These two phenomena probably interact. More stringent conservation regulations may make the licensing process even longer and, in the absence of strong monitoring, encourage illegal deforestation in the presence of high price incentives. However, this leakage effect can be contained if price incentives are controlled. Hence, our results suggest that, in the context of oil palm-imputed deforestation, a market-based instrument may help regulatory instruments be more effective.

A sector-wide tax on CPO, levied at palm oil mills and refunded against proof of sustainable production would not need local monitoring and hence not reintroduce the risk that weak institutions hinder effective forest conservation intervention (Heine et al. 2020). Indonesian Nationally Determined Contributions (NDC) to the Paris Agreement include an emission (including LUC) reduction target of 29% below business as usual by 2030. We estimate that reducing annual deforestation by 29% with respect to the 2002-2014 average could be achieved with a 19% tax on CPO. We see at least two reasons why even

greater emission reductions could be achieved with a 19% tax on CPO. First, because a tax would moderate the profitability of illegal deforestation, shrinking the leakage from legal to illegal economically motivated deforestation that our results document, and thus make regulatory conservation instruments more effective. Second, because the tax revenues could be redistributed to compensate plantations claiming (and proving) avoided deforestation, thus strengthening the price gradient between deforestation-free and deforestation-based CPO and increasing even further the incentive to avoid deforestation.

Finally, our results seem to suggest that the price incentives provided by the Roundtable on Sustainable Palm Oil (RSPO) are insufficient to reach zero-deforestation palm oil; the price premium offered by the RSPO is around 2% according to Levin (2012), and 7% according to Preusser (2015).

Further research We do not attempt in this paper to properly simulate policy effects on deforestation through prices. We do not model a separation between deforestation-free and deforestation-based markets (and prices) that is caused by a label or by downstream due diligence on sustainability. Hence, our study does not provide strong insights into the incentivizing scheme of the Roundtable on Sustainable Palm Oil (RSPO). We leave such efforts to further research.

We note that our new spatially explicit micro-economic panel dataset of palm oil mills could be useful to study the economic causes of other important phenomenon in Indonesia, like land conflicts or forest and peat fires. These data can also help further the understanding of the economics of palm oil mills, whose operations have remained a black box so far.

5.7 Appendix

Empirical framework

Estimation strategy

Functional form and estimation In this study, we estimate an exponential mean model by Poisson quasi-maximum likelihood. The Poisson distributional assumption has been made elsewhere in statistical studies of (Indonesian) deforestation (e.g., Burgess et al. (2012), Busch et al. (2012), and Busch et al. (2015)). Hence, we also seek comparability of our results with, in particular, Busch et al. (2015). The quasi-Poisson distribution imposes weaker assumptions on our data, as it only requires the mean (and not the variance) to be correctly specified. We use the standard log-link function. We perform the estimation of equation (2) with the feglm algorithm from the R package *fixest*. This method estimates generalized linear models using weighted ordinary least squares (OLS) estimations with demeaning along fixed effect dimensions in the OLS steps and no presence of the incidental parameter problem (Bergé 2018).

Partial effects

In all regressions, the price signal variable is scaled to the natural logarithm. The partial effects of price signals on deforestation are computed as the relative difference between predicted deforestation at the sample means, with and without a 1% increase in the price signal, multiplied by 100 (hence, all estimates are scaled to percentage points). From Equation 5.2, this simplifies to $100(1.01^{\hat{\alpha}} - 1)\%$ and hence does not depend on sample means (Bellavia et al. 2015). This only slightly differs from the exponential of regression coefficients as it gauges the effect for a "full" 1% change in a right-hand-side variable and not for an infinitesimal change. We present results this way because it is more consistent with computation of effects for larger changes (e.g., one standard deviation) or when second-order terms are included on the right-hand side. We estimate the variance of the partial effect with the delta method (Greene 2012).

To investigate synergies, in Section 5.5, we use interaction terms: right-hand-side variables computed as the product of the treatment (price signal here) and an interacting variable which is also featured in the right-hand side. Because our model is not linear, the informative estimate is the partial effect of the interaction term, not its coefficient (Ai and Norton 2003). Hence, interaction estimates discussed in Section 5.5 and displayed in Tables 5.6, 5.5 and 5.11 are second-order cross-derivatives of predicted deforestation, evaluated at the sample mean.

Robustness analysis

Here, we document a battery of alternative estimation and identification strategies. We explain why these different specifications are relevant and we justify why we do not keep them in our main analysis. Figure 5.1 shows how they compare with the overall price elasticity in Indonesian plantations estimated with the main specification (Equation 5.2) and sample described above. We mention only single departures from the main specification. We do not discuss combinations of alternative specifications.

IBS data cleaning We check two departures from our main analysis in terms of preparation of IBS variables, including price signals and ownership controls.

The first one is the set of imputations described in the Appendix. In our main analysis, we use stronger imputations, in order to reduce statistical noise due to duplicates and outliers, in particular. The weaker cleaning imputations are much less precise.

In the second one we check the estimate difference due to not lagging IBS variables. Recall that, in our main analysis, we lag IBS variables to correct for a suspected measurement lag between them and remote sensing variables. Not lagging IBS variables yields a lower point estimate. This does not disprove our belief that prices recorded in IBS in a given year have little effect on the deforestation recorded that same year.

Sampling We report the price elasticity estimates for two additional sampling conditions. In our main analysis, no such conditions are applied.

Under the first one, we include in the sample only plantations where more than 50% of the area was covered with primary forest in 2000. This condition is relevant because it makes the sample more homogeneous in terms of initial land use. It is not included in our main analysis because it also limits the external validity of our results.

Under the second condition, we include in the sample only the plantations for which the set of known reachable mills is constituted of at least 50% of IBS geo-localized mills. This excludes plantations for which the measurement error is too high due to our geo-localized IBS mill data set not being exhaustive. In our main analysis, we do not apply this condition for the sake of generality and simplicity.

Both conditions yield very similar estimates to the main one.

Catchment modelling How we model the true relationships between mills and plantations is a critical point in our analysis. Therefore, we explore four alternatives to the

model used in our main estimation strategy - catchment radii of 30 km in Sumatra and 50 km in Kalimantan.

The first alternative consists in the assumption that plantations are only influenced by prices at the nearest mill. This is the simplest model possible. Not surprisingly, it is very imprecise. This estimate's confidence interval is so large that we do not feature it in Figure 5.1 for the sake of readability.

The second one consists in removing from the analysis extensive margin deforestation. As defined in Section 5.2, we deem deforestation to be at the extensive margin if it occurs close[43] to a mill and prior to its establishment date. Extensive margin deforestation likely results from integrated plantation-mill decisions that are not much influenced by price signals from other mills. Indeed, the price elasticity of intensive margin deforestation is higher than that of deforestation at both extensive and intensive margins. For the sake of generality and simplicity, we consider deforestation at both margins in our main analysis.

The third alternative is a different catchment radius in each island: 50 km in Sumatra and 30 km in Kalimantan. In Section 5.3, we discuss the size of the catchment radius and the reason why it should be lower in Sumatra than in Kalimantan. The alternative catchment radii yield a higher but less precise estimate.[44] This loss of precision makes us more confident that our choice of catchment radii in the main analysis is efficient to model the relationships between plantations and mills.

Finally, we model the catchment area of each mill not as a circle defined by a radius, but as the set of plantations that can reach the mill within two hours of driving (see Harahap et al. (2019) for a discussion on the driving time[45]). This modelling is highly relevant because often, mills, although close to plantations in straight line distance, may actually not be reachable in time by trucks following weaving roads (and the opposite is also true). However, this modelling is not done in our main, preferred analysis because it may introduce endogeneity. Indeed, plantations likely expand (and hence deforest more) in parts of districts where the road infrastructure is better, while in the same area, prices are probably affected by the better access to markets enabled by better roads. This bias should be attenuated in our main analysis as we arbitrarily draw a line beyond which plantations are not connected to a mill although the road infrastructure would actually make the mill's prices influence deforestation. The estimate under this catchment area model is lower than in the main analysis and not statistically significantly different from zero. We interpret this as resulting from a negative bias due to the endogeneity introduced in catchment modelling with driving-time constraints.

[43]Deforestation is deemed close to a mill if it is occurring in one of the four nearest grid cells to this mill. Four 900ha grid cells make a 3600ha area which is enough to supply an average mill (Paoli et al. 2013).

[44]We also get an estimate under a 10 km catchment radius assumption, but here again we do not present it in Figure 5.1 as the confidence interval is so wide that it complicates the reading of the whole figure.

[45]Harahap et al. (2019) use a four-hour constraint, grounding on `https://goldenagri.com.sg/plantation-mill-24-hours/`. Here we present a twice shorter constraint because the estimation with the four-hour constraint yields too large a confidence interval to be displayed next to the other estimates.

Price signal time average As explained in Section 5.3, our main measure of price signal is a 4-year average of annual price signals. We present here price elasticity estimates with different time average lengths.

Unsurprisingly, the short-run, annual price signal measure alone yields a non-significant estimate. Indeed, we expect the development of perennial crops to have little responsiveness to annual variations. This is confirmed by the narrow confidence interval.

The price elasticity point estimate increases with the average length of the price signal time, while precision decreases. With a 5-year average, too much noise enters the price signal measure and the price elasticity becomes less precise.

Distributional assumptions Our preferred distributional assumption is a quasi-Poisson distribution (that allows the variance to be different from the mean). A Poisson distribution assumption yields the same point estimate and very similar standard errors. This suggests that our data are not subject to over or under dispersion. The negative binomial distribution assumption is another option for count data. In our case, it yields a slightly higher but less precise estimate.

Control variables We explore specifications with all combinations of control variables (except the case without any). These include the mill ownership and the number of reachable mills control variables in our main specification, and three additional control variables.

The first of these is the one-year lagged outcome variable, i.e., deforestation. Deforestation has been often shown to be an auto-regressive process, and indeed we find that, in our data, lagged deforestation is positively correlated with current deforestation (results available upon request). Furthermore, we expect that prices from the 4 past years that we average in our price signal measure also influenced past deforestation. Indeed, in our data, we find that a price signal measured as an average of prices over 3 years does influence deforestation (cf. the above paragraph on different time average lengths). However, we do not believe that plantation site level lagged deforestation can impact price signals. Therefore, we suspect lagged deforestation to be an intermediate factor. We find that neither the magnitude nor the precision of our estimate varies with the inclusion of lagged deforestation. Thus, we conclude that the effect we measure is not inflated by the spurious accumulation of intermediate effects by which past prices would cause past deforestation that would then cause present deforestation.

The second additional control variable is the (inverse-distance weighted) average share of crude palm oil (CPO) exported by reachable mills. This proxies plantation exposure to international supply chains and hence might control for additional potentially confounding systematic differences between plantations. Adding it to the main control set yields a

similar estimate.

The third additional control variable is the baseline forest trend. This is built as an interaction between the primary forest cover in 2000 and the year. It captures differential trends between plantations with different initial land uses. Adding it to the main control set yields a similar estimate.

We present estimates with a fourth control variable: the past deforestation of neighboring sites. This is measured as the average of deforestation in the 8 neighboring plantation sites (grid cells) four years ago. As such, this variable captures the potential bias that could arise from global spatial spillovers (LeSage 2014). These spillovers occur when deforestation in surrounding areas affects local deforestation. They are likely to occur (Robalino and Pfaff 2012; Shevade and Loboda 2019), and in particular it is possible that surrounding deforestation in the past, (i.e., temporally and spatially lagged) affects current local deforestation. Such spillovers can bias our estimates if past surrounding deforestation also affects current local price signals (which are 4-year averages). It could well be, indeed, that around a plantation site in a given year t-4, deforestation was important enough so that four years later, when palm trees bear their first fruits, local prices in year t are impacted.

Over all plantation and deforestation types, controlling for the neighbors' past deforestation, the price elasticity point estimate is 2.18 (confidence interval [0.58; 3.78]). This contrasts slightly with our main estimate of 1.6. This difference seems to be due to the time period over which the regression with this control variable is estimated. The long lag (4 years) in deforestation restricts the estimating sample to the time period 2005-2014 (as deforestation is observed only as of 2001). Estimating our main model over this same period yields a point estimate of 2.17 (confidence interval [0.53; 3.8]). We see at least two explanations for the absence of bias from the neighbors' past deforestation. First, the important and heterogeneous time lapse between deforestation (observed in t-4) and palm tree planting mitigates the effect of deforestation - even aggregated over 8 plantation sites - on prices. Second, the limited market power of mills on the crude palm oil market makes it less likely that deforestation - even aggregated over 8 plantation sites - affects prices.

Given the substantial change in the sampling time period implied by the addition of this control variable and its negligible incidence, we do not investigate it in combination with the other robustness control variables presented above.

Fixed effects Our main analysis uses a combination of plantation and district-year fixed effects, as we believe that most price endogeneity arises at the district level. Different fixed effects absorb variations at different levels. The plantation fixed effect only controls time-invariant heterogeneity but still allows macro-level shocks to confound the estimate, leading to less precise estimates. Adding a year fixed effect additionally controls for country-wide annual shocks, but not for more local confounding shocks. Adding, rather,

a local-year fixed effect, i.e., ruling out common shocks at the level of province, district, subdistrict or village, yields positive estimates. These are precise in the case of province-year and district-year fixed effects, larger but less precise in the case of subdistrict-year fixed effects, and very imprecise in the case of village-year fixed-effects (which we do not display in Figure 5.1 in order to better read it). This shows that most of the effect of price signals on deforestation is at play above the village-year level.

Clustering We show in Figure 5.1 how allowing correlations in standard errors within different observation clusters affects confidence intervals. Price elasticity estimates are statistically different from zero with more clusters than in our main analysis - i.e., with plantation and village clusters. They also remain significant with larger and hence fewer clusters; namely, with district clusters and two-way plantation and district-year clusters.

Appendix Tables

Table 5.8: IBS descriptive statistics

	Geo-localized IBS palm oil mills n = 587 mills			All IBS palm oil mills n = 930 mills			t-test	KS test
	mean	std.dev.	median [min; max]	mean	std.dev.	median [min; max]	p-value	p-value
First year in IBS	1999	8.19	2001 [1975; 2015]	2000	8.78	2002 [1975; 2015]	0.000	0.000
FFB farm gate price (USD/ton)	124.7	35.69	127.4 [16.84; 241.5]	123.3	35.73	125.8 [16.84; 242.2]	0.108	0.274
FFB input (ton)	149047	115114	133193 [0; 1035319]	148035	114416	132552 [0; 1035319]	0.692	1.000
CPO farm gate price (USD/ton	684.9	172.5	706.8 [170.1; 1191]	679.8	173.4	700.8 [170.1; 1191]	0.192	0.287
CPO output (ton)	36082	24384	32902 [0.64; 179142]	35795	24363	32389 [0.64; 179142]	0.587	0.999
PKO farm gate price (USD/ton)	399.9	140	389.4 [12.53; 827]	398.4	139.8	386 [12.53; 832.9]	0.676	1.000
PKO output (ton)	8441	8918	6917 [0.11; 96775]	8368	8861	6846 [0.11; 96775]	0.724	1.000
CPO export share (%)	16.85	33.37	0 [0; 100]	15.75	32.55	0 [0; 100]	0.072	0.375
Central government ownership (%)	15.39	35.48	0 [0; 100]	14.64	34.76	0 [0; 100]	0.227	0.961
Local government ownership (%)	2.25	14.65	0 [0; 100]	2.1	14.17	0 [0; 100]	0.562	1.000
National private ownership (%)	65.75	46.02	100 [0; 100]	66.76	45.7	100 [0; 100]	0.214	0.831
Foreign ownership (%)	16.62	34.89	0 [0; 100]	16.51	34.88	0 [0; 100]	0.862	1.000

Notes. This table reports summary statistics for set of variables from the Indonesian manufacturing census (IBS), at the palm oil mill level, annually in 1998-2015. The sample of geo-localized IBS palm oil mills is a sub-sample of all IBS palm oil mills. IBS palm oil mills are identified here as IBS plants that report crude palm oil (CPO) or palm kernel oil (PKO) outputs, or fresh fruit bunches (FFB) inputs at least one year, and are not in Java nor Bali islands. Farm gate prices are measured with mean unitary values (the ratios of value on quantity). USD is 2010-constant. We report p-values of Welch two-sided t-tests where the null hypothesis is that the true difference in means between the two groups is null, and the groups' variances are not assumed to be equal; and p-values of Kolmogorov-Smirnov tests where the null hypothesis is that the variables in the two groups are drawn from the same continuous distribution.

Table 5.9: Price elasticity and partial effects of control variables on deforestation across Indonesian oil palm sectors

	Industrial plantations			Smallholder plantations			All		
	Legal	Illegal	All	Legal	Illegal	All	Legal	Illegal	All
Price elasticity									
Estimate	0.51	5.18	2.11	0.14	2.09	1.53	0.2	3.03	**1.64**
95% CI	[-2.02; 3.04]	[1.58; 8.78]	[0.11; 4.11]	[-2.22; 2.5]	[0.71; 3.47]	[0.25; 2.81]	[-1.56; 1.95]	[1.16; 4.9]	**[0.3; 2.98]**
Partial effects of:									
Domestic private mill ownership									
Estimate	0.44	-0.86	-0.38	-1.21	0.36	0.57	0.32	-0.13	**-0.05**
95% CI	[-1.67; 2.55]	[-3.21; 1.48]	[-1.77; 1]	[-3.06; 0.64]	[-1.4; 2.11]	[-0.31; 1.45]	[-1.42; 2.06]	[-1.58; 1.32]	**[-1.12; 1.02]**
Foreign mill ownership									
Estimate	0.1	-2.32	-1.32	-2.78	-0.23	-0.4	0.02	-1.77	**-1.06**
95% CI	[-2.3; 2.51]	[-6.23; 1.58]	[-2.96; 0.32]	[-5.78; 0.22]	[-3.01; 2.54]	[-2.24; 1.43]	[-1.97; 2.02]	[-3.96; 0.41]	**[-2.4; 0.28]**
# reachable mills									
Estimate	-7.18	22.5	-5.44	-0.72	0.47	-2.73	-8.34	7.43	**-5.94**
95% CI	[-26.32; 11.96]	[-8.32; 53.32]	[-19.76; 8.88]	[-14.94; 13.51]	[-15.51; 16.45]	[-12.29; 6.83]	[-23.63; 6.94]	[-6.04; 20.9]	**[-15.8; 3.93]**
Observations	13081	4951	25249	2971	3412	8611	15139	7848	**31650**
Clusters	218	176	326	79	83	152	236	204	**362**

Notes. This table shows our main estimates of the price elasticity of deforestation, along with estimated partial effects of control variables. Price elasticity estimates are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The price signal is measured as the 4-year average of annual inverse-distance weighted averages of crude palm oil prices at the gates of reachable mills. Ownership variables are measured as the annual inverse-distance weighted averages of the capital shares of reachable mills. Public (either local or central) ownership is the excluded category. The number of reachable mills is measured as the count of mills within 50 km (30 km in Sumatra) annually, based on the universe of known palm oil mills. The estimates of the covariates are also scaled to percentage points. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate industrial from smallholder plantations based on scale and landscape criteria (Austin et al. 2017; Petersen et al. 2016). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

Table 5.10: Price elasticities of deforestation across the oil palm sector, by island

	Industrial plantations			Smallholder plantations			All		
	Legal	Illegal	All	Legal	Illegal	All	Legal	Illegal	All
Sumatra									
Estimate	1.98	5.51	3.09	0.17	2.09	1.54	0.37	3.03	**1.75**
95% CI	[-1.1; 5.06]	[1.45; 9.58]	[0.72; 5.46]	[-2.19; 2.53]	[0.71; 3.47]	[0.26; 2.81]	[-1.19; 1.93]	[1.07; 4.99]	**[0.34; 3.16]**
Observations	4434	3332	11680	2624	3407	8137	6183	6226	**17668**
Clusters	117	115	210	72	80	141	135	143	**247**
Kalimantan									
Estimate	-0.53	3.12	0.95	-17.86		-14.15	-0.55	3.1	**1.02**
95% CI	[-3.91; 2.85]	[-2.34; 8.58]	[-2.23; 4.12]	[-43.45; 7.72]		[-41.91; 13.6]	[-3.91; 2.82]	[-2.36; 8.56]	**[-2.12; 4.15]**
Observations	8647	1619	13569	347		474	8956	1622	**13982**
Clusters	101	61	117	7		11	102	62	**117**

Notes. This table shows our estimates of the price elasticity of deforestation by island. They are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The price signal is measured as the 4-year average of annual inverse-distance weighted averages of crude palm oil prices at the gates of reachable mills. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate industrial from smallholder plantations based on scale and landscape criteria (Austin et al. 2017; Petersen et al. 2016). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

Table 5.11: Price elasticity heterogeneity across ownership and local market development

	Industrial plantations			Smallholder plantations			All		
	Legal	Illegal	All	Legal	Illegal	All	Legal	Illegal	All
Price signal									
Estimate	1.3331	4.8232	2.4158	0.2293	1.5775	1.3196	0.9656	2.6192	1.835
95% CI	[-1.0289; 3.695]	[1.4534; 8.1929]	[0.3851; 4.4466]	[-1.7578; 2.2163]	[-0.229; 3.384]	[-0.0757; 2.715]	[-0.6907; 2.622]	[0.9758; 4.2626]	[0.4542; 3.2158]
Interaction with									
Domestic private ownership									
Estimate	8e-04	9e-04	1e-04	5e-04	4e-04	2e-04	7e-04	8e-04	2e-04
95% CI	[0; 0.0017]	[-0.0026; 0.0045]	[-6e-04; 8e-04]	[-1e-04; 0.0011]	[-7e-04; 0.0015]	[-2e-04; 7e-04]	[2e-04; 0.0012]	[-4e-04; 0.0019]	[-3e-04; 7e-04]
Foreign ownership									
Estimate	0.0012	-1e-04	3e-04	-3e-04	3e-04	-5e-04	0.001	0	3e-04
95% CI	[4e-04; 0.002]	[-0.0042; 0.004]	[-5e-04; 0.0011]	[-0.0018; 0.0012]	[-0.0015; 0.002]	[-0.0014; 5e-04]	[3e-04; 0.0017]	[-0.0015; 0.0015]	[-4e-04; 0.001]
# reachable mills									
Estimate	0.002	0.0099	-0.001	-3e-04	-0.0031	-0.0016	0.0019	0.0014	-0.0012
95% CI	[-0.0016; 0.0056]	[-0.0032; 0.023]	[-0.0049; 0.0029]	[-0.0045; 0.0039]	[-0.0104; 0.0042]	[-0.0054; 0.0021]	[-8e-04; 0.0046]	[-0.0041; 0.0069]	[-0.0038; 0.0015]
Observations	13081	4951	25249	2971	3412	8611	15139	7848	31650
Clusters	218	176	326	79	83	152	236	204	362

Notes. This table shows our estimates of the price elasticity of deforestation, along with estimated partial effects of interaction variables. Price elasticity estimates are to be interpreted as points of percentage change in average deforestation associated with a 1% increase in price signals. The price signal is measured as the 4-year average of annual inverse-distance weighted averages of crude palm oil prices at the gates of reachable mills. Interaction terms are the product of the price signal and interacting variables, or covariates. These interacting variables are ownership variables measured as the annual inverse-distance weighted averages of the capital shares of reachable mills (public ownership is the excluded category), as well as the annual count of all known reachable mills. The partial effects of interaction terms are second-order cross derivatives evaluated at the sample mean. Deforestation is measured as primary forest loss eventually replaced with oil palm plantations. We differentiate industrial from smallholder plantations based on scale and landscape criteria (Austin et al. 2017; Petersen et al. 2016). We identify illegal deforestation as occurring outside a known oil palm concession and inside a permanent forest zone designation. There are places where not enough information is available to designate the legal status. All estimates are derived from a generalized linear model of the quasi-Poisson family. All regressions include unit and district-year fixed effects, as well as ownership shares and the annual count of reachable mills as covariates. Sample observations are annual records of 3×3 km grid cells in Sumatra and Kalimantan from 2002 to 2014. They all have a positive remaining primary forest extent, and are within a 50 km (30 km in Sumatra) radius from at least one of our sample mills. 95% confidence intervals (CI) are based on standard errors computed with the delta method and clustered at the subdistrict (kecamatan) level.

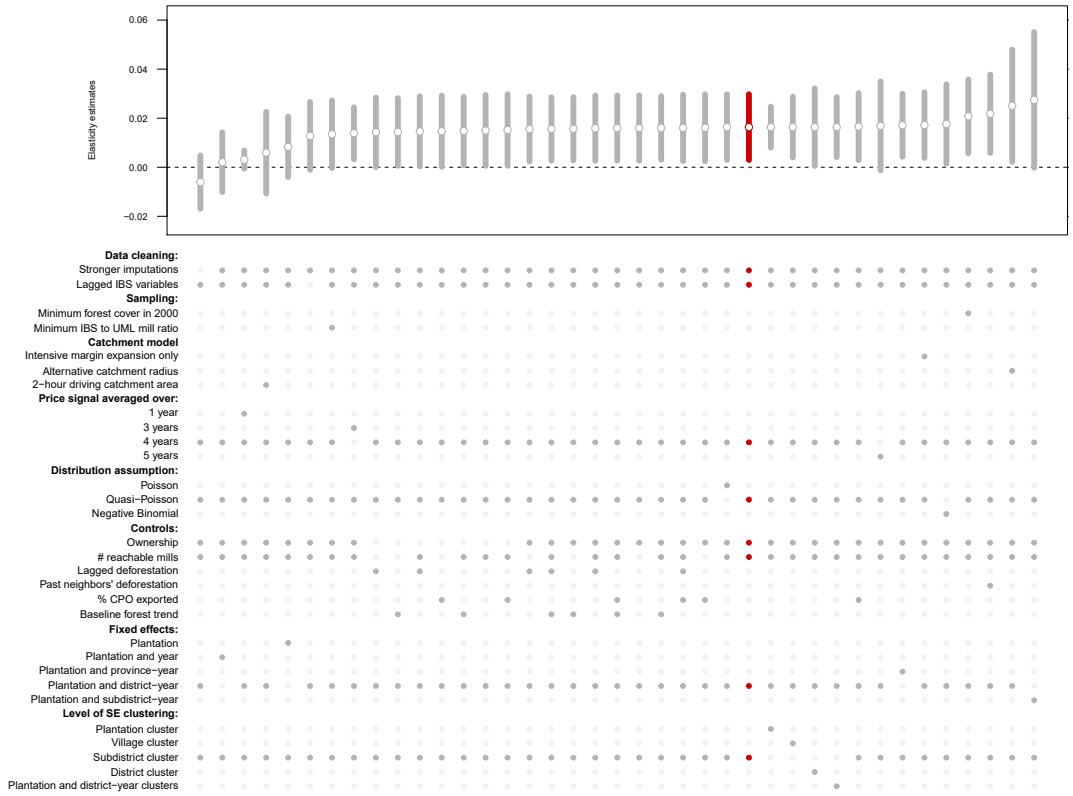
Table 5.12: p-values from equality tests of price elasticities

	All plantations			Industrial plantations	Smallholder plantations
Ho	Legal	Illegal	All		
industrial = smallholders	0.8333	0.1300	0.6536		
legal = illegal			0.0257	0.0413	0.1422

Notes. This table shows p-values of two-sided t-tests, where the null hypothesis is that the true difference in price elasticities of deforestation between two groups is null.

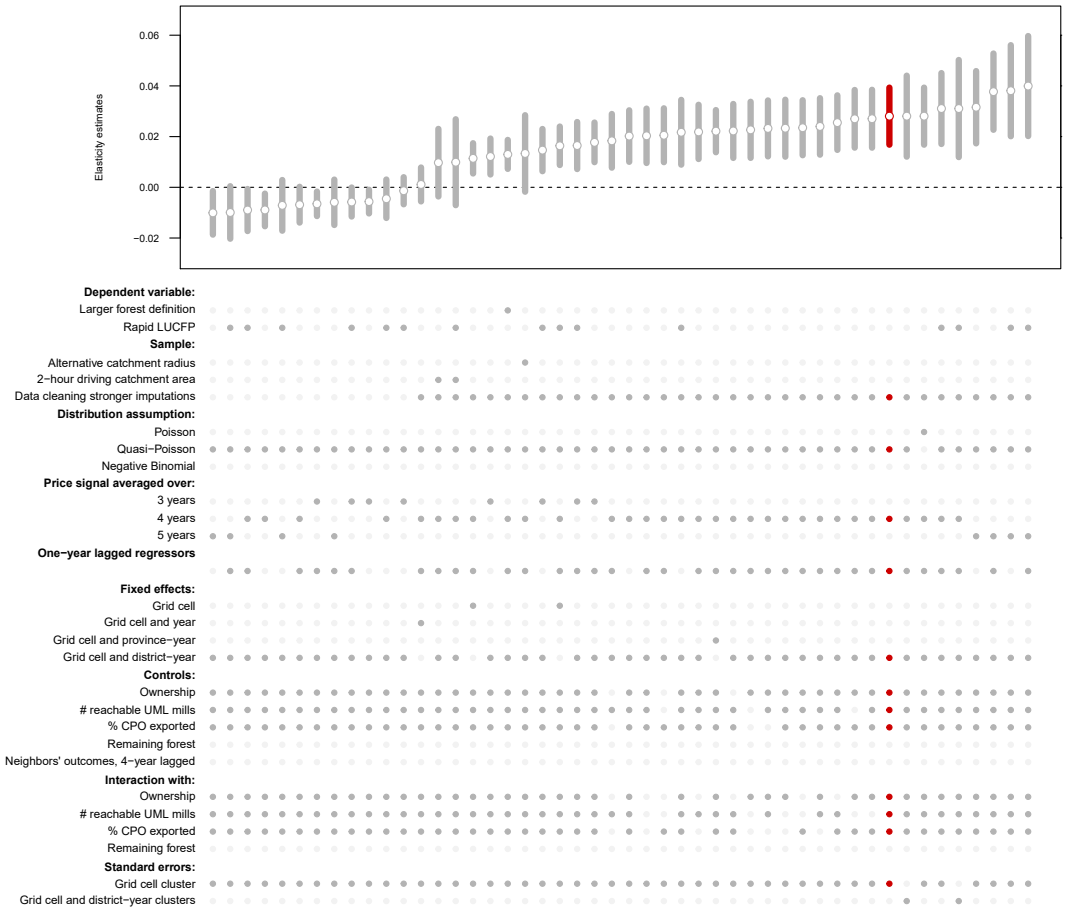
Appendix Figures

Figure 5.1: Estimates of the Indonesian price elasticity of deforestation under different specifications



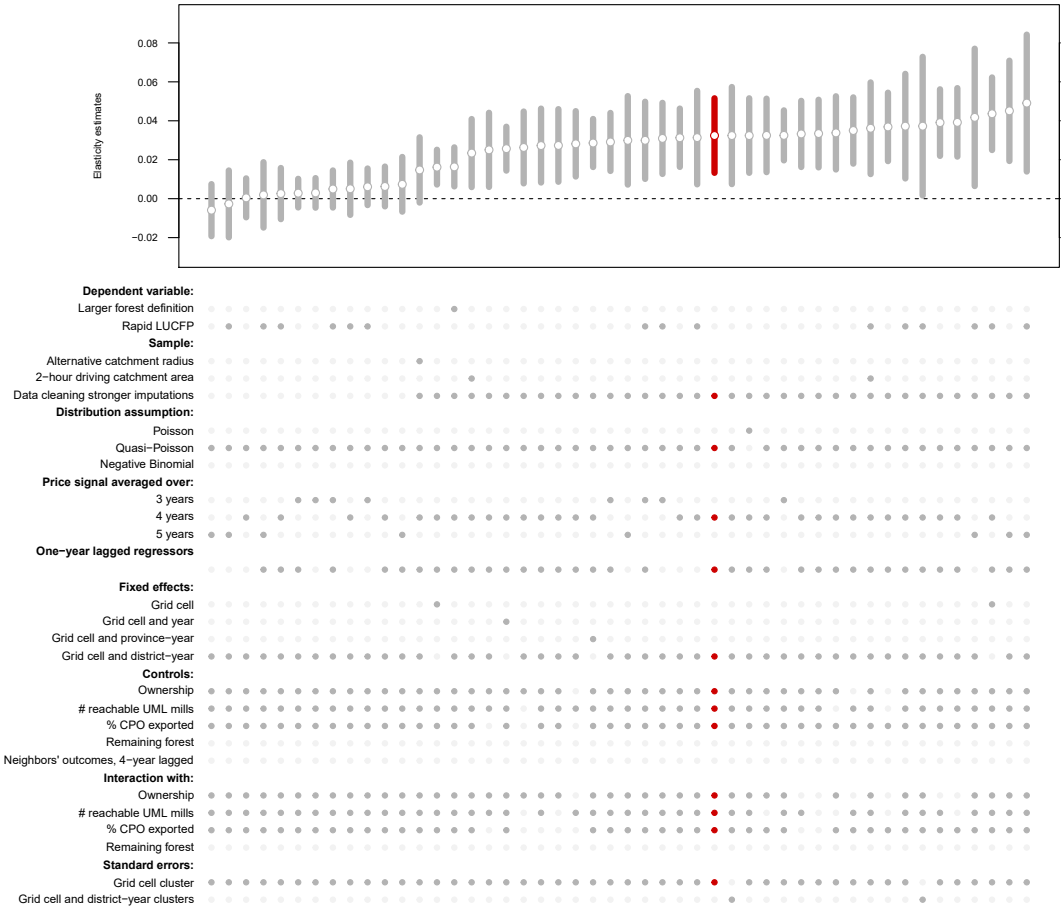
Notes. This figure shows point estimates (white dots in upper panel) of the overall Indonesian price elasticity of deforestation estimated in this paper. Grey bars in the upper panel represent 95% confidence intervals. Darker marks in the lower panel mean that the corresponding vertical estimate is derived from a model that has the corresponding horizontal feature. Our main specification is highlighted. The minimum forest cover in 2000 is 50%. The IBS to UML mill ratio designates the number of mills from our sample relative to the total number of known reachable mills. It is also set to 50% (included). Alternative catchment radius is 50 km in Sumatra and 30 km in Kalimantan.

Figure 5.2: Estimates of price elasticity of land use change from forest to industrial plantations in Sumatra and Kalimantan under different specifications



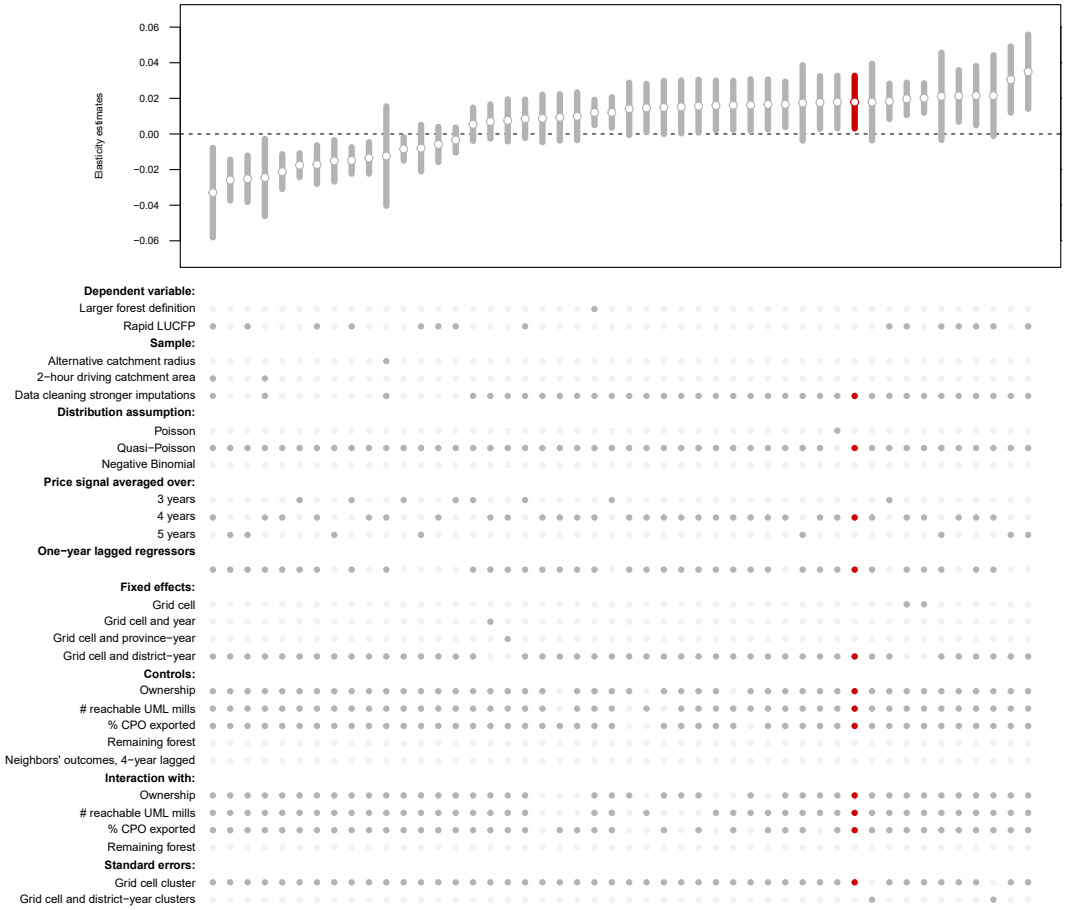
Notes. This figure shows point estimates (white dots in upper panel) of the relative change in the mean land use change from primary forest to industrial oil palm plantations (LUCFP) associated with a 1% increase in the crude palm oil (CPO) 4-year average price in Sumatra and Kalimantan. Grey bars in the upper panel represent 95% confidence intervals. Darker marks in the lower panel mean that the vertically corresponding estimate is derived from a model that has the horizontally corresponding feature. Highlighted is the main specification.
Larger forest definition refers to LUCFP occurring in >30% tree cover forest and outside 2000 industrial oil palm plantations. Alternative catchment radius is 50 km in Sumatra and 30 km in Kalimantan. Data cleaning imputations relate to the Indonesian manufacturing census data. All estimations were performed with the R package {fixest}. Standard errors were computed with the delta method.

Figure 5.3: Estimates of price elasticity of land use change from forest to industrial plantations in Sumatra under different specifications



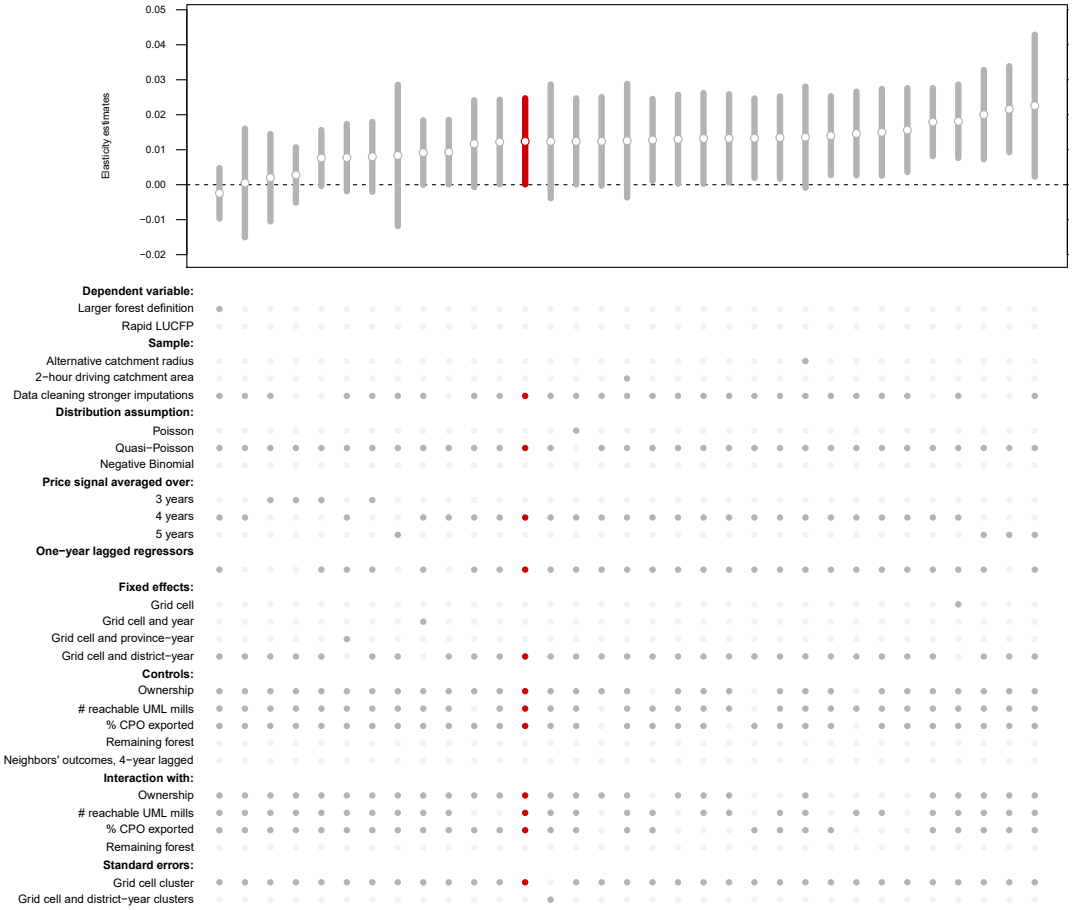
Notes. This figure shows point estimates (white dots in upper panel) of the relative change in the mean land use change from primary forest to industrial oil palm plantations (LUCFP) associated with a 1% increase in the crude palm oil (CPO) 4-year average price in Sumatra. Grey bars in the upper panel represent 95% confidence intervals. Darker marks in the lower panel mean that the vertically corresponding estimate is derived from a model that has the horizontally corresponding feature. Highlighted is the main specification.
Larger forest definition refers to LUCFP occurring in >30% tree cover forest and outside 2000 industrial oil palm plantations. Alternative catchment radius is 50 km. Data cleaning imputations relate to the Indonesian manufacturing census data. All estimations were performed with the R package {fixest}. Standard errors were computed with the delta method.

Figure 5.4: Estimates of price elasticity of land use change from forest to industrial plantations in Kalimantan under different specifications



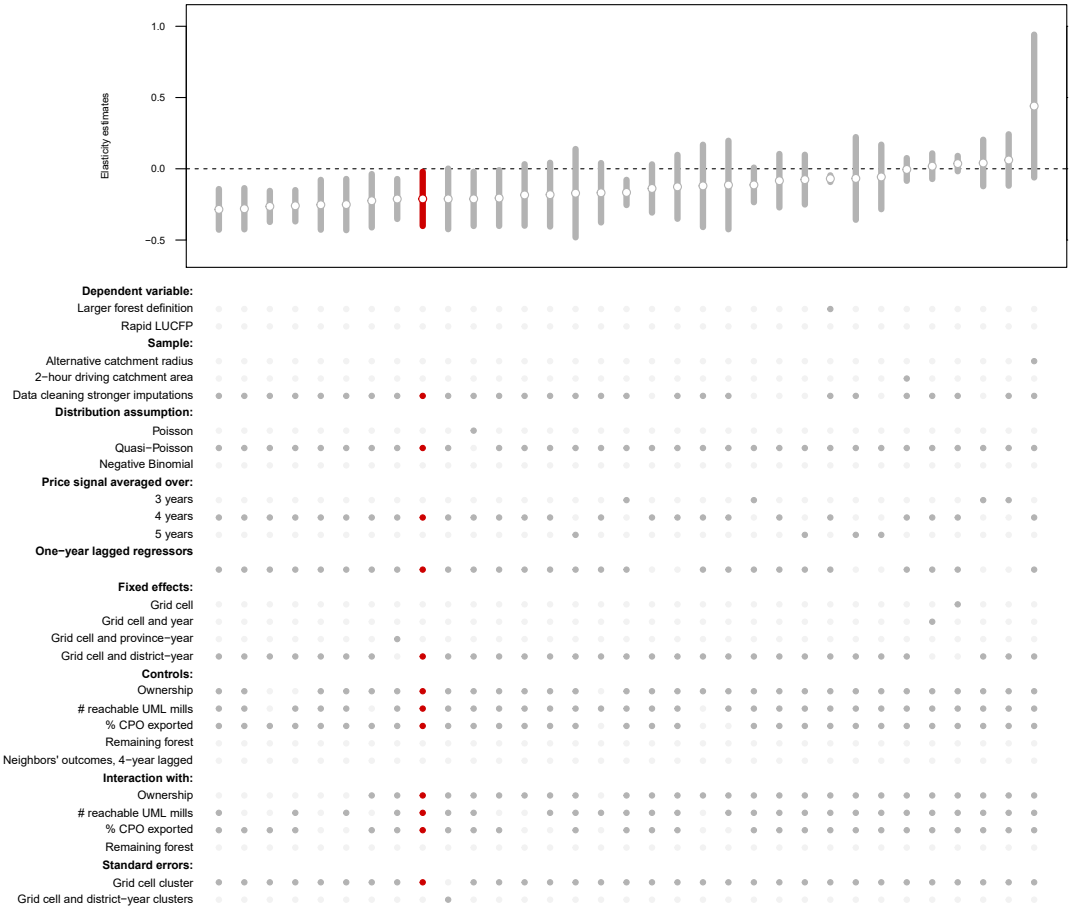
Notes. This figure shows point estimates (white dots in upper panel) of the relative change in the mean land use change from primary forest to industrial oil palm plantations (LUCFP) associated with a 1% increase in the crude palm oil (CPO) 4-year average price in Kalimantan. Grey bars in the upper panel represent 95% confidence intervals. Darker marks in the lower panel mean that the vertically corresponding estimate is derived from a model that has the horizontally corresponding feature. Highlighted is the main specification.
Larger forest definition refers to LUCFP occurring in >30% tree cover forest and outside 2000 industrial oil palm plantations. Alternative catchment radius is 30 km. Data cleaning imputations relate to the Indonesian manufacturing census data. All estimations were performed with the R package {fixest}. Standard errors were computed with the delta method.

Figure 5.5: Estimates of price elasticity of land use change from forest to smallholder plantations in Sumatra under different specifications



Notes. This figure shows point estimates (white dots in upper panel) of the relative change in the mean land use change from primary forest to smallholder oil palm plantations (LUCFP) associated with a 1% increase in the crude palm oil (CPO) 4-year average price in Sumatra. Grey bars in the upper panel represent 95% confidence intervals. Darker marks in the lower panel mean that the vertically corresponding estimate is derived from a model that has the horizontally corresponding feature. Highlighted is the main specification.
Larger forest definition refers to LUCFP occurring in >30% tree cover forest and outside 2000 industrial oil palm plantations. Alternative catchment radius is 50 km. Data cleaning imputations relate to the Indonesian manufacturing census data. All estimations were performed with the R package {fixest}. Standard errors were computed with the delta method.

Figure 5.6: Estimates of price elasticity of land use change from forest to industrial plantations in Kalimantan under different specifications



Notes. This figure shows point estimates (white dots in upper panel) of the relative change in the mean land use change from primary forest to industrial oil palm plantations (LUCFP) associated with a 1% increase in the crude palm oil (CPO) 4-year average price in Kalimantan. Grey bars in the upper panel represent 95% confidence intervals. Darker marks in the lower panel mean that the vertically corresponding estimate is derived from a model that has the horizontally corresponding feature. Highlighted is the main specification.
Larger forest definition refers to LUCFP occurring in >30% tree cover forest and outside 2000 industrial oil palm plantations. Alternative catchment radius is 30 km. Data cleaning imputations relate to the Indonesian manufacturing census data. All estimations were performed with the R package {fixest}. Standard errors were computed with the delta method.

5.8 References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge (2017). *When Should You Adjust Standard Errors for Clustering?* Working Paper 24003. National Bureau of Economic Research. DOI: 10.3386/w24003.

Ai, Chunrong and Edward C. Norton (2003). "Interaction Terms in Logit and Probit Models". *Economics Letters* 80.1, pp. 123–129. DOI: 10.1016/s0165-1765(03)00032-6.

Amiti, Mary and Jozef Konings (2007). "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia". *The American Economic Review* 97.5, pp. 1611–1638.

Austin, K.G., A. Mosnier, J. Pirker, I. McCallum, S. Fritz, and P.S. Kasibhatla (2017). "Shifting Patterns of Oil Palm Driven Deforestation in Indonesia and Implications for Zero-Deforestation Commitments". *Land Use Policy* 69, pp. 41–48. DOI: 10.1016/j.landusepol.2017.08.036.

Austin, Kemen G, Amanda Schwantes, Yaofeng Gu, and Prasad S Kasibhatla (2019). "What Causes Deforestation in Indonesia?" *Environmental Research Letters* 14.2, p. 024007. DOI: 10.1088/1748-9326/aaf6db.

Baudoin, Alice, P.M. Bosc, C Bessou, and P Levang (2017). *Review of the Diversity of Palm Oil Production Systems in Indonesia: Case Study of Two Provinces: Riau and Jambi*. Center for International Forestry Research (CIFOR). DOI: 10.17528/cifor/006462.

Bellavia, Andrea, Matteo Bottai, Andrea Discacciati, and Nicola Orsini (2015). "Adjusted Survival Curves with Multivariable Laplace Regression:" *Epidemiology* 26.2, e17–e18. DOI: 10.1097/ede.0000000000000248.

Bellemare, Marc F., Takaaki Masaki, and Thomas B. Pepinsky (2017). "Lagged Explanatory Variables and the Estimation of Causal Effect". *The Journal of Politics* 79.3, pp. 949–963. DOI: 10.1086/690946.

Bergé, Laurent (2018). *Efficient Estimation of Maximum Likelihood Models with Multiple Fixed-Effects: The R Package FENmlm*. 18-13. Department of Economics at the University of Luxembourg.

Berry, Steven T (2011). "Biofuels Policy and the Empirical Inputs to GTAP Models". Mimeo.

Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber (2012). "The Political Economy of Deforestation in the Tropics*". *The Quarterly Journal of Economics* 127.4, pp. 1707–1754. DOI: 10.1093/qje/qjs034.

Busch, J., R. N. Lubowski, F. Godoy, M. Steininger, A. A. Yusuf, K. Austin, J. Hewson, D. Juhn, M. Farid, and F. Boltz (2012). "Structuring Economic Incentives to Reduce Emissions from Deforestation within Indonesia". *Proceedings of the National Academy of Sciences* 109.4, pp. 1062–1067. DOI: 10.1073/pnas.1109034109.

Busch, Jonah and Kalifi Ferretti-Gallon (2017). "What Drives Deforestation and What Stops It? A Meta-Analysis". *Review of Environmental Economics and Policy* 11.1, pp. 3–23. DOI: 10.1093/reep/rew013.

Busch, Jonah, Kalifi Ferretti-Gallon, Jens Engelmann, Max Wright, Kemen G. Austin, Fred Stolle, Svetlana Turubanova, Peter V. Potapov, Belinda Margono, Matthew C. Hansen, and Alessandro Baccini (2015). "Reductions in Emissions from Deforestation from Indonesia's Moratorium on New Oil Palm, Timber, and Logging Concessions". *Proceedings of the National Academy of Sciences* 112.5, pp. 1328–1333. DOI: 10.1073/pnas.1412514112.

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

Cacho, Oscar J., Sarah Milne, Ricardo Gonzalez, and Luca Tacconi (2014). "Benefits and Costs of Deforestation by Smallholders: Implications for Forest Conservation and Climate Policy". *Ecological Economics* 107, pp. 321–332. DOI: 10.1016/j.ecolecon.2014.09.012.

Carlson, Kimberly M., Robert Heilmayr, Holly K. Gibbs, Praveen Noojipady, David N. Burns, Douglas C. Morton, Nathalie F. Walker, Gary D. Paoli, and Claire Kremen (2018). "Effect of Oil Palm Sustainability Certification on Deforestation and Fire in Indonesia". *Proceedings of the National Academy of Sciences* 115.1, pp. 121–126. DOI: 10.1073/pnas.1704728114.

Cisneros, Elías, Krisztina Kis-Katos, and Nunung Nuryartono (2020). "Palm Oil and the Politics of Deforestation in Indonesia". *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3547436.

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

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. DOI: 10.1093/restud/rdz017.

Euler, Michael, Stefan Schwarze, Hermanto Siregar, and Matin Qaim (2016). "Oil Palm Expansion among Smallholder Farmers in Sumatra, Indonesia". *Journal of Agricultural Economics* 67.3, pp. 658–676. DOI: 10.1111/1477-9552.12163.

Gaveau, David L. A., Douglas Sheil, Husnayaen, Mohammad A. Salim, Sanjiwana Arjasakusuma, Marc Ancrenaz, Pablo Pacheco, and Erik Meijaard (2016). "Rapid Conversions and Avoided Deforestation: Examining Four Decades of Industrial Plantation Expansion in Borneo". *Scientific Reports* 6.1. DOI: 10.1038/srep32017.

Gaveau, David L.A., Bruno Locatelli, Mohammad A. Salim, Husna Yaen, Pablo Pacheco, and Douglas Sheil (2018). "Rise and Fall of Forest Loss and Industrial Plantations in Borneo (2000-2017)". *Conservation Letters*, e12622. DOI: 10.1111/conl.12622.

Gaveau, David, Bruno Locatelli, Mohammad Salim, Husnayaen Husnayaen, Timer Manurung, Adrià Descals, Arild Angelsen, Erik Meijaard, and Douglas Sheil (2021). *Slowing Deforestation in Indonesia Follows Declining Oil Palm Expansion and Lower Oil Prices*. Preprint. DOI: 10.21203/rs.3.rs-143515/v1.

GoI (2016). "First Nationally Determined Contribution Republic Of Indonesia".

Goldman, Elizabeth, Mikaela J Weisse, Nancy Harris, and Martina Schneider (2020). "Estimating The Role Of Seven Commodities In Agriculture-linked Deforestation: Oil Palm, Soy, Cattle, Wood Fiber, Cocoa, Coffee, And Rubber". Mimeo.

Greene, W. H. (2012). *Econometric Analysis*. 7th ed. Prentice Hall: Upper Saddle River, NJ.

Greenpeace (2011). "Indonesia Ministry of Forestry, Greenpeace, and WRI. "Indonesia Oil Palm Concessions." Accessed through Global Forest Watch in October 2020. Www.Globalforestwatch.Org."

Guillaume, Thomas, Martyna M. Kotowska, Dietrich Hertel, Alexander Knohl, Valentyna Krashevska, Kukuh Murtilaksono, Stefan Scheu, and Yakov Kuzyakov (2018). "Carbon Costs and Benefits of Indonesian Rainforest Conversion to Plantations". *Nature Communications* 9.1, p. 2388. DOI: 10.1038/s41467-018-04755-y.

Gunarso, Petrus, Manjela Eko Hartoyo, Fahmuddin Agus, and Timothy J Killeen (2013). "Oil Palm and Land Use Change in Indonesia, Malaysia and Papua New Guinea". Mimeo.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend (2013). "High-Resolution Global Maps of 21st-Century Forest Cover Change". *Science* 342.6160, pp. 850–853. DOI: 10.1126/science.1244693.

Hansen, M., P. Potapov, B. Margono, S. Stehman, S. Turubanova, and A. Tyukavina (2014). "Response to Comment on "High-Resolution Global Maps of 21st-Century Forest Cover Change"". *Science* 344.6187, pp. 981–981. DOI: 10.1126/science.1248817.

Harahap, Fumi, Sylvain Leduc, Sennai Mesfun, Dilip Khatiwada, Florian Kraxner, and Semida Silveira (2019). "Opportunities to Optimize the Palm Oil Supply Chain in Sumatra, Indonesia". *Energies* 12.3, p. 420. DOI: 10.3390/en12030420.

Harris, Nancy L, Kevin Brown, Michael Netzer, and Petrus Gunarso (2013). "Projections of Oil Palm Expansion in Indonesia, Malaysia and Papua New Guinea from 2010 to 2050". Mimeo.

Heilmayr, Robert, Kimberly M. Carlson, and Jason Jon Benedict (2020). "Deforestation Spillovers from Oil Palm Sustainability Certification". *Environmental Research Letters*. DOI: 10.1088/1748-9326/ab7f0c.

Heine, Dirk, Erin Hayde, and Michael Faure (2020). "Letting Commodity Tax Rates Vary With the Sustainability of Production", p. 47.

Hertel, Thomas W (2018). "Economic Perspectives on Land Use Change and Leakage". *Environmental Research Letters* 13.7, p. 075012. DOI: 10.1088/1748-9326/aad2a4.

Hsiao, Allan (2020). "Coordination and Commitment in International Climate Action: Evidence from Palm Oil". Mimeo.

Hughes, Alice C. (2018). "Have Indo-Malaysian Forests Reached the End of the Road?" *Biological Conservation* 223, pp. 129–137. DOI: 10.1016/j.biocon.2018.04.029.

Jelsma, Idsert, G. C. Schoneveld, Annelies Zoomers, and A. C. M. van Westen (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. DOI: 10.1016/j.landusepol.2017.08.012.

Kharina, Anastasia, Chris Malins, and Stephanie Searle (2016). "Biofuels Policy In Indonesia: Overview And Status Report". Mimeo.

Khatiwada, Dilip, Carl Palmén, and Semida Silveira (2018). "Evaluating the Palm Oil Demand in Indonesia: Production Trends, Yields, and Emerging Issues". *Biofuels*, pp. 1–13. DOI: 10.1080/17597269.2018.1461520.

Krishna, Vijesh V., Christoph Kubitza, Unai Pascual, and Matin Qaim (2017). "Land Markets, Property Rights, and Deforestation: Insights from Indonesia". *World Development* 99, pp. 335–349. DOI: 10.1016/j.worlddev.2017.05.018.

Leblois, Antoine, Olivier Damette, and Julien Wolfersberger (2017). "What Has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data". *World Development* 92, pp. 82–102. DOI: 10.1016/j.worlddev.2016.11.012.

Lee, Janice Ser Huay, Sinan Abood, Jaboury Ghazoul, Baba Barus, Krystof Obidzinski, and Lian Pin Koh (2014). "Environmental Impacts of Large-Scale Oil Palm Enterprises Exceed That of Smallholdings in Indonesia: Forest Loss from Sumatra's Oil Palm Industry". *Conservation Letters* 7.1, pp. 25–33. DOI: 10.1111/conl.12039.

LeSage, James P. (2014). "What Regional Scientists Need to Know About Spatial Econometrics". *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2420725.

Levin, J (2012). *Sustainability and Profitability in the Palm Oil Sector*. WWF, FMO, CDC.

Margono, Belinda Arunarwati, Peter V. Potapov, Svetlana Turubanova, Fred Stolle, and Matthew C. Hansen (2014). "Primary Forest Cover Loss in Indonesia over 2000–2012". *Nature Climate Change* 4.8, pp. 730–735. DOI: 10.1038/nclimate2277.

Maryadi, Yusuf, A. K., and A. Mulyana (2004). "Pricing of Palm Oil Fresch Fruit Bunches for Smallholders in South Sumatra".

Masliani, M. Muslich Mustadjab, Syafrial, and Ratya Anindita (2014). "Price Determination of Palm Oil Fresh Fruit Bunches on Imperfect Competition Market in Central Kalimantan Province, Indonesia". *Journal of Economics and Sustainable Development* 5.1, pp. 134–139–139.

MoA (2010). *Area and Production by Category of Producers: Oil Palm, 1967-2010 (Indonesian Ministry of Agriculture, Jakarta, Indonesia)*.

MoF (2008). *Reducing Emissions from Deforestation and Forest Degradation in Indonesia. Indonesian Forest Climate Alliance Consolidation Report*. Ministry of Forestry of the Republic of Indonesia.

– (2019). *Kawasan Hutan 2019 - Kementerian Lingkungan Hidup Dan Kehutanan Republik Indonesia*.

Mummolo, Jonathan and Erik Peterson (2018). "Improving the Interpretation of Fixed Effects Regression Results". *Political Science Research and Methods* 6.4, pp. 829–835. DOI: 10.1017/psrm.2017.44.

Pacheco, P, S Gnych, A Dermawan, and B Okarda (2017). *The Palm Oil Global Value Chain: Implications for Economic Growth and Socialand Environmental Sustainability*. Center for International Forestry Research (CIFOR). DOI: 10.17528/cifor/006405.

Paoli, Gary D., Piers Gillespie, P.L. Wells, L. Hovani, A.E. Sileuw, N. Franklin, and Jim Schweithelm (2013). *Oil Palm in Indonesia: Governance, Decision Making and Implications for Sustainable Development*. Jakarta, Indonesia: The Nature Conservancy.

Petersen, Rachael, Dmitry Aksenov, Elena Esipova, Elizabeth Goldman, Nancy Harris, Irina Kurakina, Tatiana Loboda, Alexander Manisha, Sarah Sargent, and Varada Shevade (2016). "Mapping Tree Plantations With Multispectral Imagery: Preliminary Results For Seven Tropical Countries". Mimeo.

Pirard, Romain, S Gnych, P Pacheco, and S Lawry (2015). *Zero-Deforestation Commitments in Indonesia: Governance Challenges*. Center for International Forestry Research (CIFOR). DOI: 10.17528/cifor/005871.

Pirard, Romain, Nils Schulz, Jason Benedict, Robert Heilmayr, Ramada Febrian, Ben Ayre, and Helen Bellfield (2020). "Corporate Ownership and Dominance of Indonesia's Palm Oil Supply Chains". Mimeo.

Pirker, Johannes, Aline Mosnier, Florian Kraxner, Petr Havlík, and Michael Obersteiner (2016). "What Are the Limits to Oil Palm Expansion?" *Global Environmental Change* 40, pp. 73–81. DOI: 10.1016/j.gloenvcha.2016.06.007.

Potapov, Peter, Aleksey Yaroshenko, Svetlana Turubanova, Maxim Dubinin, Lars Laestadius, Christoph Thies, Dmitry Aksenov, Aleksey Egorov, Yelena Yesipova, Igor Glushkov, Mikhail Karpachevskiy, Anna Kostikova, Alexander Manisha, Ekaterina Tsybikova, and Ilona Zhuravleva (2008). "Mapping the World's Intact Forest Landscapes by Remote Sensing". *Ecology and Society* 13.2. DOI: 10.5751/es-02670-130251.

Preusser, S (2015). *The Correlation between Economic and Financial Viability with Sustainability for Palm Oil Plantations*. RSPO online.

Purnomo, Herry, Beni Okarda, Ade Ayu Dewayani, Made Ali, Ramadhani Achdiawan, Hariadi Kartodihardjo, Pablo Pacheco, and Kartika S. Juniwaty (2018). "Reducing Forest and Land Fires through Good Palm Oil Value Chain Governance". *Forest Policy and Economics* 91, pp. 94–106. DOI: 10.1016/j.forpol.2017.12.014.

Rahman, Ayat K Ab, Ramli Abdullah, N Balu, and Mohd Shariff (2013). "The Impact of La Niña and El Niño Events on Crude Palm Oil Prices: An Econometric Analysis". Mimeo.

Rifin, Amzul (2014). "The Effect of Progressive Export Tax on Indonesian Palm Oil Industry". Mimeo.

Robalino, Juan A. and Alexander Pfaff (2012). "Contagious Development: Neighbor Interactions in Deforestation". *Journal of Development Economics* 97.2, pp. 427–436. DOI: 10.1016/j.jdeveco.2011.06.003.

Sanders, D. J., J. V. Balagtas, and G. Gruere (2014). "Revisiting the Palm Oil Boom in South-East Asia: Fuel versus Food Demand Drivers". *Applied Economics* 46.2, pp. 127–138. DOI: 10.1080/00036846.2013.835479.

Santeramo, Fabio Gaetano and Stephanie Searle (2019). "Linking Soy Oil Demand from the US Renewable Fuel Standard to Palm Oil Expansion through an Analysis on Vegetable Oil Price Elasticities". *Energy Policy* 127, pp. 19–23. DOI: 10.1016/j.enpol.2018.11.054.

Shevade, Varada S. and Tatiana V. Loboda (2019). "Oil Palm Plantations in Peninsular Malaysia: Determinants and Constraints on Expansion". *Plos One* 14.2. Ed. by Gopalasamy Reuben Clements, e0210628. DOI: 10.1371/journal.pone.0210628.

Souza Rodrigues, E. (2019). "Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis". *The Review of Economic Studies Limited* 86.6, pp. 2713–2744.

Stavins, Robert N (1999). "The Costs of Carbon Sequestration: A Revealed-Preference Approach". *American Economic Review* 89.4, pp. 994–1009. DOI: 10.1257/aer.89.4.994.

UML (2018). *World Resources Institute, Rainforest Alliance, Proforest, and Daemeter. "Universal Mill List". October 2018. Accessed through Global Forest Watch on 01/2020.* https:www.globalforestwatch.org. URL: https:www.globalforestwatch.org (visited on 01/05/2020).

Wheeler, David, Dan Hammer, Robin Kraft, Susmita Dasgupta, and Brian Blankespoor (2013). "Economic Dynamics and Forest Clearing: A Spatial Econometric Analysis for Indonesia". *Ecological Economics*, p. 12.

Wicke, Birka, Pita Verweij, Hans van Meijl, Detlef P van Vuuren, and Andre PC Faaij (2012). "Indirect Land Use Change: Review of Existing Models and Strategies for Mitigation". *Biofuels* 3.1, pp. 87–100. DOI: 10.4155/bfs.11.154.

Woittiez, Lotte S., Mark T. van Wijk, Maja Slingerland, Meine van Noordwijk, and Ken E. Giller (2017). "Yield Gaps in Oil Palm: A Quantitative Review of Contributing Factors". *European Journal of Agronomy* 83, pp. 57–77. DOI: 10.1016/j.eja.2016.11.002.

Wooldridge, Jeffrey M. (1999). "Distribution-Free Estimation of Some Nonlinear Panel Data Models". *Journal of Econometrics* 90.1, pp. 77–97. DOI: 10.1016/s0304-4076(98)00033-5.

Zant, Wouter, Christopher L. Gilbert, and Hidde P. Smit (2004). *Feasibility of Making Price Risk Management Instruments Available to Oil Palm Smallholders in Indonesia and Thailand. Prepared for the Commodity Risk Management Group of the World Bank*. Economic and Social Institute, Free University, De Bolelaan 1105, 1081 HV Amsterdam, The Netherlands.

Zikri, Muhammad (2009). "An Econometric Model for Deforestation in Indonesia". Mimeo.

Chapter 6

Cross-country evidence on the effect of air pollution on COVID-19 from thermal inversions as a natural experiment[†,‡]

Abstract

Early COVID-19 outbreaks in Wuhan, China and Lombardy, Italy coincided with high levels of air pollution drawing attention to a potential role of air pollutants as a driver of the new respiratory disease. In this paper, we provide cross-country evidence for robust short-term effects of air pollution on COVID-19 outcomes across space and time. Using random variation in air pollution generated by thermal inversions as a natural experiment and district-level panel data from Belgium, Brazil, Germany, Italy, Great Britain, and the US, we rule out that spurious correlations from changes in mobility and economic activity are driving the results. We find that a 1%-increase in pollution leads to 5.1% more COVID-19 deaths. This increase in mortality is caused by pollution exposure around the time of infection and during the pre-symptomatic phase. We find no evidence for effects in later stages of the disease. The higher death toll is not simply driven by an increase in the total number of patients. Instead, a higher relative effect on the number of deaths than on the number of cases suggests that pollution in the early course of the disease decreases the rate of survival in registered COVID-19 patients. These results indicate that short-term measures to reduce air pollution can help mitigate the virus' spread to and severity in new patients.

6.1 Introduction

The detrimental effects of air pollutants on human health are well documented and some reductions in human exposure to them can be achieved fast and at low-cost (Landrigan et al. 2018). Since COVID-19 is a respiratory disease, policy-makers are considering short-term measures against air pollution to mitigate health damages from the disease. A UK House of Commons working group, the *All Party Parliamentary Group Air Pollution*, launched a strategy to reduce COVID-19 risks associated with air pollution (APPGAQ 2020). In France, local politicians have called on the national government to take "emergency measures", for instance to create low emissions zones, limit the use of liquid manure, and replace old wood stoves (Le Monde 2020). This study provides the first cross-country evidence for the link between contemporaneous air pollution and COVID-19 outcomes across space and time that complements previous country-level and long-term exposure studies.

[†]Hannah Klauber, Nicolas Koch, and Sebastian Kraus (2021). "Cross-country evidence on the effect of air pollution on COVID-19 from thermal inversions as a natural experiment". Under Review at *Environmental Research Letters*

[‡]We thank Jeffrey M. Wooldridge for methodological advice on the two-stage estimation procedure applied in this paper. We thank Yujie Wang for advice on the air pollution data analyzed in this paper.

Particularly for the lifting-phases of COVID-19 lockdowns the results could be relevant, since increased mobility and especially car use lead to increases in air pollution, which can partially jeopardize the achieved mitigation of infections and the severity of COVID-19.

The primary empirical challenge when estimating the effect of air pollution on COVID-19 outcomes is that air pollution is not randomly assigned to places, but results from different dimensions of human behavior, some of which cannot be measured. Air pollution levels have fallen because of COVID-19 outbreaks and lockdowns (Chen et al. 2020; He et al. 2020). This pollution reduction can be due to behavior changes as a reaction to the health threat (e.g. less car use) or due to a third, omitted variable that is driving both air pollution and COVID-19 outcomes (e.g. policy changes that occur independently from local infection clusters). We expect the pollution reductions to lead to substantial improvements in human health (Chen et al. 2020; Cicala et al. 2020) and several public health studies demonstrate an association between COVID-19 outcomes and long-term (Wu et al. 2020) and short-term (Cole et al. 2020; Isphording and Pestel 2020; Ogen 2020; Persico and Johnson 2020; Son et al. 2020-11-20, 2020; Travaglio et al. 2021; Zhu et al. 2020) air pollution. However, they are typically confined to narrow geographical areas and the fragmentary nature of the still emerging evidence makes it difficult to draw systematic inferences on the COVID-19 response to air pollution across space and time. A robust cross-country analysis has been missing.

We address this gap by providing quasi-experimental evidence based on a geographically diverse set of countries – Belgium, Brazil, Germany, Great Britain, Italy and the US. Therefore, we expect our analysis to have higher external validity than individual-country studies. By studying the timing and magnitude of the pollution effect on both COVID-19 cases and deaths, we also provide evidence on the effect mechanisms. To estimate the links between short-term increases in air pollution and COVID-19 outcomes, we use thermal inversions as a natural experiment, a well-proven strategy in the epidemiological (Anderson 2009; Firket 1931) and economic (Arceo et al. 2016; Bondy et al. 2020; Hicks et al. 2016; Jans et al. 2018; Sager 2019) air pollution literature. While temperature usually decreases with altitude leading to the rise and dispersion of air pollutants, during inversion episodes the atmospheric temperature layers are inverted and the warmer air at higher altitudes traps the pollutants close to the ground. Satellite imagery and climate models let us measure thermal inversions and air pollution at high spatial resolution. We compile data from all countries that publish COVID-19 outcomes at the district-level as of the time of the analysis to match the scale of our local weather-driven natural experiment. Our results suggest that pollution-induced COVID-19 deaths stem from exposure in the early course of the disease that leads not only to more registered COVID-19 patients but also to a lower survival rate among them.

6.2 Methods

Data

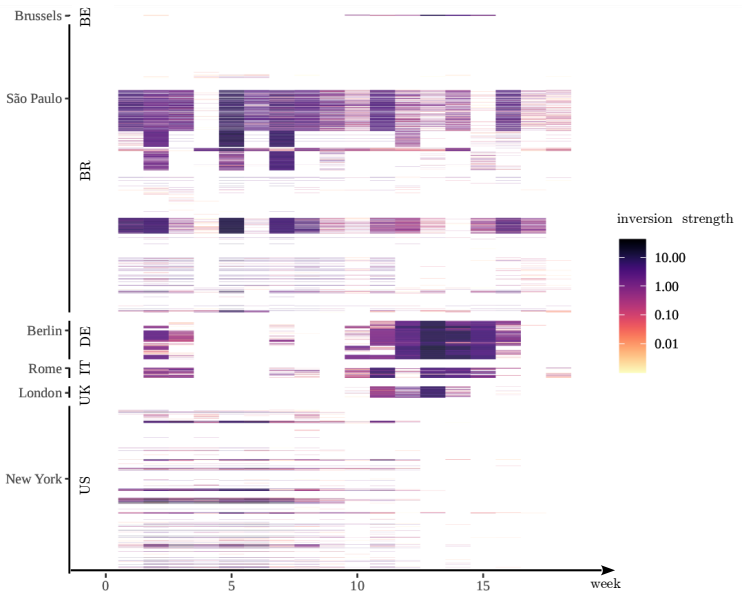
COVID-19 data District-level data on daily numbers of accumulated COVID-19 cases and deaths from official sources are available as open data online. At the time of data collection, we were able to find information at a fine spatial and temporal resolution for six countries: Belgium, Brazil, Germany, Italy, Great Britain, and the United States. We scrape all reported data for the month January through April in 2020. However, data become available at different times across countries and while we observe case numbers for all six of them, death counts are available only for Belgium,[1] Brazil, Germany and the United States (see Table 6.6 in the Appendix). Also, the reporting differs across and within countries. For instance, in some districts case numbers include only lab-confirmed cases, while in others they also include probable cases. Moreover, the time of reporting as well as the testing regimes may differ across districts and over time. While we are not able to observe these information, our empirical strategy accounts for these differences by including „fixed effects" in the regression models (see Section 6.2).

Weather and pollution data Data on atmospheric air temperature and weather conditions are from the European Centre for Medium-Range Weather Forecasts (ECMWF). The product ERA5 provides interpolated hourly data at 37 different pressure levels with a spatial resolution of 31 km (0.28125 degrees). We define inversions as episodes when the difference in temperature between the 925 hPa pressure level (ca. 600 m above sea level) and the 1,000 hPa pressure level (ca. 30 m above sea-level) is positive and weight them by their strength, i.e. the continuous difference between the two temperature layers. Following recent studies, we consider night-time inversions using temperature measurings at two a.m. local time (Jans et al. 2018; Sager 2019)[2] and aggregate the data to the district level. Figure 6.1 shows the temporal and spatial variation in the occurrence of inversion episodes across districts at the weekly level. From ERA5 we also compile data on total precipitation, UV radiation, specific humidity, air temperature two metres above the surface and U and V wind components.

We use Aerosol Optical Depth (AOD) derived from MODIS satellite-images to proxy for $PM_{2.5}$ concentration in the atmosphere. We use the MCD19A2 V6 data product provided by NASA (Lyapustin and Wang 2018) which measures AOD over land at a one kilometer resolution with global coverage. For all districts we extract daily mean values of AOD pixels whose centroid lies within the district. The Appendix 6.5 provides additional information on the weather and pollution data.

[1] Data for Belgium contain death counts only for Brussels.
[2] Day-time inversions can sometimes be seen and might affect how people behave.

Figure 6.1: Spatial and temporal variation in the occurrence of inversion episodes



Notes. The figure illustrates the frequency and strength of inversion episodes. The vertical axis refers to the districts. Countries and capital cities are labeled. The horizontal axis refers to the weeks. The color scale on the right indicates the strength of the inversions. Weeks without inversion episodes are in white.

Data on policies and mobility We use data from the Oxford COVID-19 Government Response Tracker (OxCGRT) (Hale et al. 2020), which systematically collects information on countries' policy responses to the pandemic. For our analysis we compose two control sets reflecting the extent of government action in each country on a daily level (see Appendix 6.5 for a detailed overview). The first one comprises containment and closure policies, such as closings of schools and universities. The second one comprises health system policies, such as regulations on access to PCR testing. While these data are available only at the country level, we account for policy differences at the sub-national level by including „fixed effects" in the regression models (see Section 6.2). Moreover, we use daily movement data from the social media platform Facebook, which compiles information on people's locations from their mobile devices for states and districts (Facebook 2020).

Empirical Strategy

Short-term air pollution changes could be caused by local economic conditions and human behavior that are highly correlated to COVID-19 outcomes and are barely controllable. Therefore, we aim at approximating an experimental setting where the exposure to air pollution is randomized. We exclusively rely on short-term pollution changes induced by thermal inversions, i.e. meteorological phenomena that occur independently of COVID-19 outcomes based on conditions that are well understood (Arceo et al. 2016; Bondy et al. 2020; Hicks et al. 2016; Jans et al. 2018; Sager 2019). To this end, we implement an „instrumental

variable“ strategy which in a first stage identifies air pollution changes caused by thermal inversions and in a second stage estimates the effect of the inversion-driven pollution changes on COVID-19 case and death counts. In doing so we consider inversion episodes and air pollution changes over several weeks since COVID-19 outcomes are affected with time-lags. Precisely, we aggregate over three weeks when considering case numbers and over four weeks when considering death counts. We ensure that our model only forms comparisons of observations within specified spatial and temporal units that are assumed comparable by the inclusion of „fixed effects“.

The first-stage regression is given as

$$P_{iw} = \alpha_1 I_{iw} + \alpha_2 C_{i\overline{w}_{-1}} + W'_{iw}\gamma_1 + M'_{iw}\gamma_2 + H'_{iw}\gamma_3 + \sigma_i + \tau_w + \eta_{cm} + \epsilon_{iw} \tag{6.1}$$

where the dependent variable is the logged average $PM_{2.5}$ concentration in district i over the three or four weeks preceding the current week w. The parameter of interest, α_1, represents the effect of inversions I_{iw} that are weighted by their intensity and summed up over the same three or four weeks. We control for weather W'_{iw}, COVID-19-related containment and closure measures M'_{iw}, and health system policies H'_{iw} in the same period.[3] The variable $C_{i\overline{w}_{-1}}$ is the accumulated COVID-19 outcome in the preceding week.[4] The fixed effects σ_i, τ_w, and η_{cm} account for determinants of the dependent variable specific to each district i, week w and month m in country c. For instance, they rule out that differences between districts in baseline conditions due to a history of inversions and human sorting behavior as a reaction to pollution drive the effects. The second stage includes the predicted $PM_{2.5}$ from Equation 6.1 as an explanatory variable:

$$C_{i\overline{w}_0} = \beta_1 \widehat{P_{iw}} + \beta_2 C_{i\overline{w}_{-1}} + W'_{iw}\overline{\delta_1} + M'_{iw}\overline{\delta_2} + H'_{iw}\overline{\delta_3} + \sigma_i + \tau_w + \eta_{cm} + \mu_{iw} \tag{6.2}$$

where $C_{i\overline{w}_0}$ is the accumulated COVID-19 outcome of district i in the current week. The coefficient β_1 represents the percentage change in the accumulated cases or deaths that occurred over the past week due to a 1%-increase in inversion-driven air pollution $\widehat{P_{iw}}$ over the past three or four weeks. We estimate the second stage using Poisson pseudo-maximum likelihood estimation (Correia et al. 2020) and bootstrap standard errors clustered at the district level.

To robustify our two-stage regression procedure at the weekly level, we estimate a refined model at the daily level. Note that we now measure the direct relationship (“reduced form”) between thermal inversions and COVID-19 outcomes. This is a simplified approach of estimating the effect via air pollution that leads to approximately equivalent results:

$$C_{id} = \sum_{\theta} \rho_\theta I^{\theta}_{id} + \rho_1 C_{id-1} + W'_{id}\kappa + \lambda_{im} + \zeta_{sw} + \kappa_{cd} + \nu_{id} \tag{6.3}$$

[3] Appendix 6.5 provides a detailed overview of the used control variables.

[4] This variable is included at this stage because it is required in the second stage and both regressions must include the same covariates (Wooldridge 2010).

where the COVID-19 outcomes, C_{id}, and inversion variables, I_{id}, are given for district i and day d. We include θ lags of the inversion variable, where θ runs from -20 to 0 for regressions with cases and from -27 to 0 for regressions with deaths as the dependent variable. Fixed effects λ_{im}, ζ_{sw}, and κ_{cd} absorb variation specific to a district-month, state-week, and country-day. The higher temporal resolution reduces the threat that changes in policies, economic activity, medical system performance, and testing regimes introduce trends in our model. Because our policy controls vary at the country-day level, they are redundant in this specification, and we include weather covariates only.

Our empirical strategy crucially depends on the assumption that, conditional on included variables and fixed effects, there is no other causal channel from inversions to COVID-19 outcomes. Inversions do not impact human health directly, but during the day, they can sometimes be seen (Sager 2019), which may affect how people behave. To account for this, we rely on night-time inversions only. Moreover, we conduct several robustness checks to ensure that inversions do not affect mobility patterns (Section 6.3).

6.3 Results

Thermal inversion effects on air pollution

First, we present evidence on the relationship between thermal inversions and air pollution. Table 6.1 shows the effect of an additional degree in inversion strength on the logged mean $PM_{2.5}$ concentration according to Equation 6.1. Because we assume that pollution affects infections and deaths with varying time lags, the left panel of Table 6.1 presents the inversion effect on average $PM_{2.5}$ concentrations in a three-week time window, while the right panel uses a four-week time window.

Positive and statistically significant coefficients, as well as F-statistics well above ten, robustly show that inversions increase ambient $PM_{2.5}$ pollution. For instance, column (3) indicates that a one-degree increase in the inversion strength in the three-week time window increases the average $PM_{2.5}$ concentration by about 0.58%. On average, the strength of an inversion on a single day equals 2.27. Our estimate therefore implies that an additional weekday with an inversion causes a 3.95%-increase in the weekly $PM_{2.5}$ level ($0.0058 \cdot 2.27 \cdot 3$). In the sample underlying the right panel of the table, the strength of an average inversion is 2.67. Thus, we derive a similar 3.42%-increase from column (6) ($0.0032 \cdot 2.67 \cdot 4$). These results are consistent with evidence from other studies. Inversion effects on $PM_{2.5}$ and PM_{10} in Mexico, the United States, and Sweden range from 2 to 4% when converted to the week-level (Arceo et al. 2016; Hicks et al. 2016; Jans et al. 2018). Sager (2019) shows that a one-degree change in inversion strength increases the daily $PM_{2.5}$ concentration by about 10.47% in the United Kingdom. This corresponds to a 0.50%-increase in the three-week and a 0.37%-increase in the four-week pollution

concentration, which is close to the effects in Table 6.1.

Table 6.1: The effect of thermal inversions on ambient air pollution

	logged $PM_{2.5}$					
	three-week window			four-week window		
	(1)	(2)	(3)	(4)	(5)	(6)
inversion strength	0.0053***	0.0074***	0.0058***	0.0026***	0.0031***	0.0032***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
F-statistic	93.51	163.95	104.39	17.08	23.08	25.61
Observations	101,046	101,046	101,046	65,569	65,569	65,569
Countries	BEL, BRA, DEU, GBR, ITA, USA			BEL, BRA, DEU, USAm		
weather controls	yes	yes	yes	yes	yes	yes
containment controls		yes	yes		yes	yes
health system controls			yes			yes

Notes. The table reports regressions coefficients from six estimations of Equation 6.1. In each panel control variables are added sequentially from left to right. The first set of controls contains weather variables only, the second set adds controls for COVID-related containment and closure policies (e.g. school closings and stay at home requirements), and the third set adds COVID-related health system policies (e.g. testing policies and contact tracing). All regressions include district, week and country-month fixed effects. Standard-errors clustered at the district level are in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Air pollution effects on COVID-19

Next, we estimate how inversion-driven increases in air pollution affect COVID-19 infections and deaths according to Equation 6.2. The coefficients in Table 6.2 represent elasticities, i.e., percentage changes in the dependent variable linked to a 1%-increase in the average $PM_{2.5}$ concentration over the preceding three or four weeks. For instance, column (3) indicates that with every 1%-increase in $PM_{2.5}$ over the preceding three weeks, the case numbers grow by 1.478%. Given that we control for the accumulated case numbers of the preceding week, this 1.478%-change takes place over a seven-day period. Table 6.2 also points to significant mortality effects. Column (6) shows that a 1%-higher pollution level in the past four weeks leads to 5.120% higher death counts. Again, this change refers to the death counts seven days ago.

The different coefficient magnitudes in the left and right panel, indicate a change in the survival rate of registered COVID-19 patients. If deaths increased exclusively because of an increasing number of patients, we would expect relative effects of similar magnitude. The differences in the magnitudes could be explained if air pollution increases the severity of the disease or the share of vulnerable people in registered patients. To ensure that the differences are not simply driven by the differing country samples, we re-estimate the left panel regressions including only Belgium, Brazil, Germany, and the USA, and arrive at similar results.

Table 6.2: The effect of ambient air pollution on COVID-19 outcomes

	COVID-19 cases three-week window			**COVID-19 deaths** four-week window		
	(1)	(2)	(3)	(4)	(5)	(6)
predicted logged $PM_{2.5}$	−0.023	1.018***	1.478***	4.015	4.577***	5.120***
	(0.317)	(0.189)	(0.276)	(2.607)	(1.741)	(1.632)
Observations	72,021	72,021	72,021	20,658	20,658	20,658
Countries	BEL, BRA, DEU, GBR, ITA, USA			BEL, BRA, DEU, USA		
weather controls	yes	yes	yes	yes	yes	yes
containment controls		yes	yes		yes	yes
health system controls			yes			yes

Notes. The table reports regression coefficients from estimations of Equation 6.2. In each panel control variables are added sequentially from left to right. The first set of controls contains weather variables only, the second set adds controls for COVID-related containment and closure policies (e.g. school closings and stay at home requirements), and the third set additionally adds COVID-related health system policies (e.g. testing policies and contact tracing). In the left panel we control for the accumulated case number of the preceding week, in the right panel we control for accumulated death number of the preceding week. All regressions include district, week and country-month fixed effects. Standard-errors clustered at the district level are in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

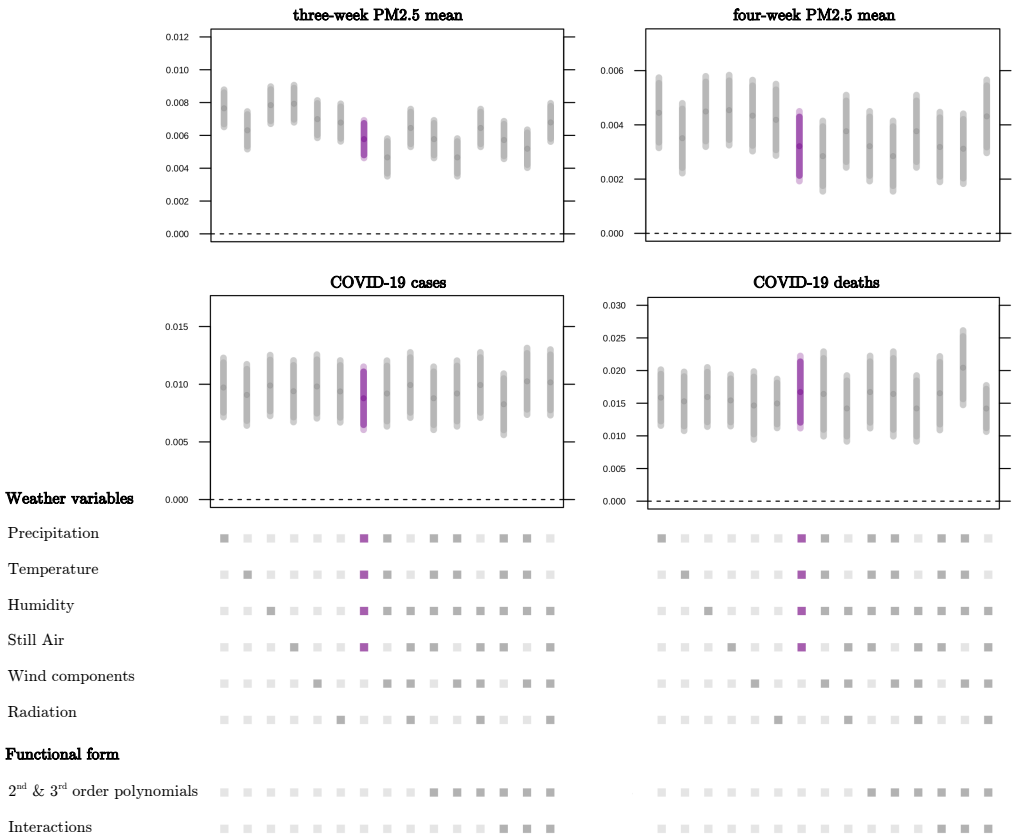
Accounting for population mobility

We conduct indirect tests to assess the assumption that the occurrence of thermal inversions does not affect COVID-19 outcomes through other channels than air pollution ("exclusion restriction"). People could be more prone to using their cars or spend more time with others indoors because of the temperature changes that create the inversions. Therefore, we analyze the effect of inversions on relative changes in movement and the time spent at different types of locations using daily movement data from the social media platform Facebook. However, the results in Table 6.7 in the Appendix (Section 6.5) do not point to causal effects of inversions on people's activities. Moreover, we expand our control set of weather covariates. Local weather conditions strongly correlate with inversions and with how and where people spend their time. Figure 6.2 demonstrates that our main results are robust across varying sets of weather controls.

Another assumption of our empirical strategy is that air pollution effects on COVID-19 outcomes do not spill across district borders. However, this may not hold when infected people travel freely. Inversions may not only affect COVID-19 outcomes in the same district but also in the surrounding ones. In this case, the estimated effects would be underestimated. To assess the extent of a potential bias, we restrict our sample to districts and weeks that are subject to a restrictive lock-down where people are required to stay at home with exceptions for daily exercise, grocery shopping, and essential trips. As these containment measures are specifically targeted to prevent virus transmission, we assume that spillover effects are less of a concern in this sample.

Table 6.3 provides regression estimates of the direct relationship between thermal inversions and COVID-19 outcomes, a simplified approach of estimating the effect via air

Figure 6.2: Alternative weather controls



Notes. The four panels in the figure present the effect of thermal inversions on the dependent variable stated above each panel. In each panel the coefficients are estimated using different weather control sets which are specified by the specification chart below. All regressions include containment and health system controls. Confidence intervals are plotted for the 5 and 10% level of statistical significance.

pollution leading to approximately equivalent results.[5] Compared to the inversion effect on COVID-19 case numbers for the full sample (0.0088) the estimated coefficient for the lock-down sample is slightly larger in magnitude (0.0134). For COVID-19 deaths, the coefficients are similar in magnitude. We conclude that if spillover effects exist, they do not bias our results markedly.

The timing of the effects on COVID-19

From the findings at the weekly level, we cannot infer when exactly the effects on COVID-19 outcomes occur after pollution increases and whether this timing is plausible given existing

[5]The effect of inversions on COVID-19 outcomes is approximately equal to the product of the effect of inversion-driven air pollution on COVID-19 outcomes (Table 6.2) and the effect of inversions on air pollution (Table 6.1). For instance, when multiplying the coefficient in column (3) in Table 6.1 with the corresponding coefficient in Table 6.2, we obtain a value of 0.0086 which is close to the estimate in column (1) in Table 6.3.

Table 6.3: The effect of thermal inversions on COVID-19 outcomes – full sample vs. lock-down sample

	COVID-19 cases three-week window		**COVID-19 deaths** four-week window	
	(1)	(2)	(3)	(4)
inversion strength	0.0088***	0.0134***	0.0167***	0.0158***
	(0.0014)	(0.0013)	(0.0028)	(0.0030)
Lock-down sample	no	yes	no	yes
Observations	72,021	13,885	20,658	5,145
Countries	BEL, DEU, GBR, ITA, USA		BEL, DEU, USA	

Notes. In the left panel we control for the accumulated case numbers of the preceding week, in the right panel we control for accumulated death number of the preceding week. All regressions include district, week and country-month fixed effects as well as weather controls. Standard-errors clustered at the district level are in parentheses. We only control for weather covariates, because the districts remaining in this sub-sample are very similar in terms of containment strategies and control variables for political measures are therefore highly collinear. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

knowledge about the disease. For this, we turn to an analysis at the daily level.

Table 6.4 reports regression coefficients from the estimation of Equation 6.3 using the total number of COVID-19 cases as the dependent variable. The coefficients indicate how the presence of a thermal inversion on the past 21 days affects cases. For instance, the first coefficient in column (1) represents how a 1-degree increase in inversion strength on the current day changes registered cases compared to those registered yesterday. The last coefficient in column (3) represents the respective effect of the same change in inversion strength 20 days ago. In turning to the daily level, we adopt more restrictive fixed effects that control for district-month, state-week and country-day specific changes in the virus development. Overall, the table indicates statistically significant increases in COVID-19 cases with a delay of five to 16 days. Note that these findings may indicate a time range when pollution effects occur. However, they should not be interpreted as indicative of specific days on which the effects occur.

To compare the effects in Table 6.4 and Table 6.2, we conduct a back-of-the envelope calculation. Dividing the sum of all 21 coefficients in Table 6.4 by the effect of a one-degree increase in inversion strength on the daily $PM_{2.5}$ mean (0.006/0.0315 = 0.1904762%) and taking it to the power of seven ($1.001904762^7 = 1.01341$), we obtain an elasticity of 1.341, which is close to the earlier estimated 1.478.

We also estimate Equation 6.3 using COVID-19 deaths as the dependent variable and considering a 28-day lag structure. Table 6.5 shows statistically significant coefficients for the 22nd to 26th lag. While the effects concentrate on the fourth week after exposure, we also observe an increase at the 16th lag. We do not observe statistically significant coefficients in the first weeks that indicate contemporaneous pollution effects on COVID-19 deaths. To compare the magnitude of the effects with those in Table 6.2, we again derive the approximate elasticity. Dividing the sum of all 28 coefficients by the daily pollution effect (0.0249/0.0315 = 0.7904762) and taking it to the power of seven ($1.007904762^7 = 1.056663$),

Table 6.4: Pollution effects on COVID-19 case numbers by days

		COVID-19 cases		
		(1) $week_{t=-1}$	(2) $week_{t=-2}$	(3) $week_{t=-3}$
inversion strength	$day_{t=1}$	−0.0000	0.0003	0.0012***
		(0.0004)	(0.0004)	(0.0004)
inversion strength	$day_{t=2}$	0.0003	0.0008*	0.0006*
		(0.0005)	(0.0005)	(0.0003)
inversion strength	$day_{t=3}$	0.0000	0.0004	−0.0008**
		(0.0005)	(0.0007)	(0.0003)
inversion strength	$day_{t=4}$	0.0007	0.0010*	−0.0006*
		(0.0005)	(0.0007)	(0.0003)
inversion strength	$day_{t=5}$	0.0011*	0.0002	0.0001
		(0.0006)	(0.0005)	(0.0003)
inversion strength	$day_{t=6}$	0.0003	0.0002	−0.0000
		(0.0006)	(0.0004)	(0.0003)
inversion strength	$day_{t=7}$	−0.0002	0.0000	0.0004
		(0.0005)	(0.0005)	(0.0003)
Countries		BEL, BRA, DEU, GBR, ITA, USA		

Notes. The table reports regression coefficients from an estimation of Equation 6.3 using COVID-19 case numbers as the dependent variable. The regression includes district-month, state-week and country-day fixed effects. We control for weather covariates as well as accumulated case numbers of the preceding day. Standard-errors clustered at the district level are in parentheses. The sample size is 236,074. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

we obtain an elasticity of 5.667. This number deviates only slightly from the earlier estimated 5.120.

Table 6.5: Pollution effects on COVID-19 deaths by days

		COVID-19 deaths			
		(1) $week_{t=-1}$	(2) $week_{t=-2}$	(3) $week_{t=-3}$	(4) $week_{t=-4}$
inversion strength	$day_{t=1}$	0.0019	0.0007	0.0006	0.0027***
		(0.0019)	(0.0014)	(0.0012)	(0.0010)
inversion strength	$day_{t=2}$	0.0004	0.0014	0.0030**	0.0021**
		(0.0024)	(0.0015)	(0.0012)	(0.0009)
inversion strength	$day_{t=3}$	0.0005	−0.0016	0.0010	0.0016*
		(0.0025)	(0.0014)	(0.0011)	(0.0008)
inversion strength	$day_{t=4}$	−0.0020	0.0007	−0.0002	0.0015
		(0.0023)	(0.0016)	(0.0012)	(0.0009)
inversion strength	$day_{t=5}$	0.0003	0.0020	0.0003	0.0025***
		(0.0018)	(0.0016)	(0.0011)	(0.0009)
inversion strength	$day_{t=6}$	0.0005	0.0009	0.0016	0.0016
		(0.0019)	(0.0014)	(0.0010)	(0.0010)
inversion strength	$day_{t=7}$	−0.0012	0.0001	0.0010	0.0010
		(0.0018)	(0.0015)	(0.0011)	(0.0008)
Countries		BEL, BRA, DEU, USA			

Notes. The table reports regression coefficients from an estimation of Equation 6.3 using COVID-19 death numbers as the dependent variable. The regression includes district-month, state-week and country-day fixed effects. We control for weather covariates as well as accumulated death numbers of the preceding day. Standard-errors clustered at the district level are in parentheses. The sample size is 82,423. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1<

6.4 Discussion and conclusion

The revealed patterns in the daily analysis suggest increases in cases and deaths that are consistent with common priors on time lags of the disease and its measurements. With regard to the case numbers, we observe no significant effects in the first days after exposure, but only five to 16 days later. In fact, if symptom development or the transmission probability was affected by air pollution, changes in case numbers are unlikely to occur in the days immediately after the pollution incident, because the average incubation time equals five days (Lauer et al. 2020) and testing and reporting introduce an additional time lag (Abbott et al. 2020). Moreover, we would expect positive effects roughly in a time frame of two weeks after the pollution incident, as the majority of infected people develops symptoms within 12 days (Lauer et al. 2020). Different effect mechanisms may explain the observed pattern. $PM_{2.5}$ exposure is linked to cardiovascular and respiratory morbidity (Landrigan et al. 2018) which is associated with worse COVID-19 outcomes (Yang et al. 2020). By affecting the cardiovascular and respiratory system, short-term pollution exposure may also cause otherwise asymptomatic COVID-19 cases to become symptomatic. Moreover, air pollution hampers the immune response to infections (Becker and Soukup 1999; Ciencewicki and Jaspers 2007; Lambert et al. 2003; Rivellese and Prediletto 2020) and propagates their transmission (Chen et al. 2010; Peng et al. 2020; Ye et al. 2016).

With regard to the death numbers, we find significant effects with a delay of 22 to 26 days, a pattern that matches the one obeserved for the case numbers. This is because the median number of days between first symptoms and death is 18 (Zhou et al. 2020). If air pollution on the current day leads already infected people to develop symptoms and newly infected to develop symptoms after the incubation time of about five days, we would expect to see effects mainly around the 18th and the 23rd lag or somewhat later because of administrative delays. We do not observe effects on the death numbers in earlier lags that could point to a worsening of a COVID-19 infection in patients who are already in a critical condition.

Overall, the findings highlight an important role of pollution exposure around the time of infection and during the pre-symptomatic phase. Moreover, the higher relative effect on the number of deaths than on the number of cases suggests that pollution does not only increase COVID-19 patient numbers but decreases the rate of survival among them. These findings are consistent with experimental studies (Becker and Soukup 1999; Ciencewicki and Jaspers 2007; Lambert et al. 2003) suggesting that pollution exposure suppresses early immune responses to infections and consequently leads to more severe inflammation and worse prognosis in the later course of a disease. Our findings indicate that even short-term reductions in air pollution can help mitigate the spread and severity of COVID-19. Since pollution reductions are known to generate large net benefits, particularly in countries with high pollution levels, many short-term measures to curb air pollution are low regrets options, as long as they do not divert attention from the core measures needed to mitigate

COVID-19 directly.

We only estimate the net short-term effect of air pollution on COVID-19 outcomes, controlling for baseline differences between populations in mid- and long-run exposure. Therefore, it can be considered a lower bound for the potential benefits from reducing air pollution over longer time periods. Estimates of the heterogeneity in COVID-19 outcomes in terms of mid- and long-term air pollution exposure are complementary to our approach and could help target policies at vulnerable populations, particularly if analyzed in interaction with short-term air pollution shocks. Moreover, our approach does not speak to the biological mechanisms creating the effect. To learn more about these underlying mechanisms, combining quasi-experimental variation in air pollution with individual-level health observations might be helpful. Differences in air filtration in the workplace or schools created by regulations or renovation cycles could serve as a suitable analysis setting.

6.5 Appendix

Data

COVID-19 data

We compile daily data on COVID-19 cases and deaths at the district level. The following table provides an overview of the available data across countries. Data availability differs across countries, e.g. Germany started reporting on January 15th, while Brazil started reporting on February 25th, 2020.

Table 6.6: Overview of the COVID-19 data sources

Country	Cases	Deaths	Original Source	First reported	Provided by
Belgium	yes	partly	Sciensano	22-01-2020	Corona Data Scraper (2020)
Brazil	yes	yes	Ministério da Saúde	25-02-2020	Cota (2020)
Germany	yes	yes	Robert Koch-Institut	15-01-2020	Robert Koch Institut (2020)
Italy	yes	no	Dept. of Civil Protection	24-02-2020	Krispin (2020)
United Kingdom	yes	no	Dept. of Health and Social Care	30-01-2020	UK Government (2020)
United States	yes	yes	local governments and health departments	21-01-2020	The New York Times (2020)

Weather and pollution data We obtain data on atmospheric air temperature as well as weather conditions from ECMWF (The European Centre for Medium-Range Weather Forecasts). The product ERA5 provides interpolated hourly data at 37 different pressure levels with a spatial resolution of 31 km (0.28125 degrees) which we obtain for the months January through April in 2020. We aggregate the gridded ERA5 data to the district level

(GADM-GID2) by calculating the weighted mean, where the weights equal the fraction of the grid covered by the district. Following He et al. (2020), we construct an indicator for still air from the wind component data, which is equal to one for wind speed no greater than one m/s. For our analysis at the weekly level, we sum up all inversions within moving three-week and four week time windows.

We use Aerosol Optical Depth (AOD) derived from MODIS satellite-images to proxy for $PM_{2.5}$ concentration in the atmosphere. We use the MCD19A2 V6 data product provided by NASA (Lyapustin and Wang 2018) which measures AOD (blue band 0.47 µm) over land at a one kilometer resolution with global coverage combining images from the Terra and Aqua satellites. The data set can be used on Google Earth Engine.

We extract daily mean values of AOD pixels whose centroid lies in one of the districts, for which we have COVID-19 data. Note that Google Earth Engine's $.reduceRegions$ function does not extract values from pixels that do not have their centroid in the polygon of interest and therefore only works well with high resolution data, if study areas are small. For most longitudes both the Terra and the Aqua satellite pass over at around the same time during the day with a difference of around ±1.5 hours. Grid cells at latitudes higher than 50 degrees tend to have several overpasses per day. The time attribute on the images is registered in UTC. Since we have study areas from different continents, we convert the time attribute into local solar time to avoid systematic differences in the time of the day at which we measure AOD. We only use images with the best quality according to the AOD_QA band excluding for instance measurements from areas with cloud cover.

For our regression analysis we aggregate the daily observations to the week level and average over three weeks for regressions explaining COVID-19 cases and over four weeks for deaths. The aggregation accounts for potentially lagged effects of inversions on air quality.

Control variables

Weather covariates In all of our main regressions we include the variables precipitation, humidity, temperature, and still air. However, we test 14 alternative sets of weather covariates including additional variables on wind and radiation and allowing for interactions between the variables and higher order polynomials (see Figure 6.2).

Closure and containment policies This set of control variables comprises school and work-place closings. When analyzing our larger regression sample with COVID-19 cases as the dependent variable, we additionally control for closings of public transport, cancellations of public events, orders to "shelter-in-place" and otherwise confine to the home, and restrictions on internal movement between cities or regions. To prevent high

degrees of collinearity, these variables are not included when we analyze the smaller sample with COVID-19 deaths as the dependent variable.

Health system policies This set accounts for regulations on access to PCR testing. When analyzing our larger regression sample with COVID-19 cases as the dependent variable, we additionally control for the strategies for contact tracing that are executed after a positive diagnosis. For the same reason mentioned above, the analysis of pollution effects on COVID-19 deaths does not include this control variable.

Accounting for population mobility

Table 6.7: The effect of thermal inversions on movement

	Relative change in movement	Proportion staying within a single location
	(1)	(2)
inversion strength	0.0010	−0.0000
	(0.0007)	(0.0003)
Countries	BEL, BRA,DEU, GBR, ITA, USA	

Notes. For every country we use data at the highest resolution available, i.e. district-level data for Brazil and the United States and state-level data for the European countries in our sample. The regression includes district-month, state-week and country-day fixed effects. We control for weather covariates. Standard-errors clustered at the district/state level are in parentheses. The sample size is 276,418. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

6.6 References

Abbott, S, J Hellewell, RN Thompson, K Sherratt, HP Gibbs, NI Bosse, JD Munday, S Meakin, EL Doughty, JY Chun, YWD Chan, F Finger, P Campbell, A Endo, CAB Pearson, A Gimma, T Russell, null null, S Flasche, AJ Kucharski, RM Eggo, and S Funk (2020). "Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using National and Subnational Case Counts". *Wellcome Open Research* 5.112. DOI: 10.12688/wellcomeopenres.16006.1.

Anderson, H.R. (2009). "Air Pollution and Mortality: A History". *Atmospheric Environment* 43.1, pp. 142–152. DOI: 10.1016/j.atmosenv.2008.09.026.

APPGAQ (2020). *Air Quality Strategy to Reduce Coronavirus Infection*. All Party Parliamentary Group Air Pollution.

Arceo, Eva, Rema Hanna, and Paulina Oliva (2016). "Does the Effect of Pollution on Infant Mortality Differ between Developing and Developed Countries? Evidence from Mexico City". *The Economic Journal* 126.591, pp. 257–280.

Becker, Susanne and Joleen M Soukup (1999). "Exposure to Urban Air Particulates Alters the Macrophage-Mediated Inflammatory Response to Respiratory Viral Infection". *Journal of Toxicology and Environmental Health Part A* 57.7, pp. 445–457.

Bondy, Malvina, Sefi Roth, and Lutz Sager (2020). "Crime Is in the Air: The Contemporaneous Relationship between Air Pollution and Crime". *Journal of the Association of Environmental and Resource Economists* 7.3, pp. 555–585.

Chen, Kai, Meng Wang, Conghong Huang, Patrick L Kinney, and Paul T Anastas (2020). "Air Pollution Reduction and Mortality Benefit during the COVID-19 Outbreak in China". *The Lancet Planetary Health* 4.6, e210–e212.

Chen, Pei-Shih, Feng Ta Tsai, Chien Kun Lin, Chun-Yuh Yang, Chang-Chuan Chan, Chea-Yuan Young, and Chien-Hung Lee (2010). "Ambient Influenza and Avian Influenza Virus during Dust Storm Days and Background Days". *Environmental health perspectives* 118.9, pp. 1211–1216.

Cicala, Steve, Stephen P Holland, Erin T Mansur, Nicholas Z Muller, and Andrew J Yates (2020). *Expected Health Effects of Reduced Air Pollution from COVID-19 Social Distancing*. Working Paper 27135. National Bureau of Economic Research. DOI: 10.3386/w27135.

Ciencewicki, Jonathan and Ilona Jaspers (2007). "Air Pollution and Respiratory Viral Infection". *Inhalation toxicology* 19.14, pp. 1135–1146.

Cole, M.A., C. Ozgen, and E. Strobl (2020). "Air Pollution Exposure and Covid-19 in Dutch Municipalities". *Environmental and Resource Economics* 76.4, pp. 581–610.

Corona Data Scraper (2020). URL: https://coronadatascraper.com/#home (visited on 06/09/2020).

Correia, Sergio, Paulo Guimarães, and Tom Zylkin (2020). "Fast Poisson Estimation with High-Dimensional Fixed Effects". *The Stata Journal* 20.1, pp. 95–115. DOI: 10.1177/1536867X20909691.

Cota, Wesley (2020). "Monitoring the Number of COVID-19 Cases and Deaths in Brazil at Municipal and Federative Units Level". *SciELOPreprints:362*. DOI: 10.1590/scielopreprints.362.

Facebook (2020). *Protecting Privacy in Facebook Mobility Data during the COVID-19 Response*. URL: https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/ (visited on 07/29/2020).

Firket, Jean (1931). "Sur Les Causes Des Accidents Survenus Dans La Vallée de La Meuse, Lors Des Brouillards de Décembre 1930". *Bulletin de l'Academie royale de medecine de Belgique* 11.5, pp. 683–741.

Hale, Thomas, Sam Webster, Anna Petherick, Toby Phillips, and Beatriz Kira (2020). *Oxford COVID-19 Government Response Tracker*. URL: https://github.com/OxCGRT/covid-policy-tracker (visited on 08/01/2020).

He, Guojun, Yuhang Pan, and Takanao Tanaka (2020). "The Short-Term Impacts of COVID-19 Lockdown on Urban Air Pollution in China". *Nature Sustainability*, pp. 1–7. DOI: 10.1038/s41893-020-0581-y.

Hicks, Daniel, Patrick Marsh, and Paulina Oliva (2016). "Air Pollution and Procyclical Mortality: Causal Evidence from Thermal Inversions". Mimeo.

Isphording, Ingo E. and Nico Pestel (2020). "Pandemic Meets Pollution: Poor Air Quality Increases Deaths by Covid-19". *Journal of Environmental Economics and Management (forthcoming)* (ID 3643182).

Jans, Jenny, Per Johansson, and J Peter Nilsson (2018). "Economic Status, Air Quality, and Child Health: Evidence from Inversion Episodes". *Journal of Health economics* 61, pp. 220–232.

Krispin, Rami (2020). *Covid19italy*. URL: https://github.com/RamiKrispin/covid19Italy (visited on 06/09/2020).

Lambert, Amy L, James B Mangum, Michael P DeLorme, and Jeffrey I Everitt (2003). "Ultrafine Carbon Black Particles Enhance Respiratory Syncytial Virus-Induced Airway Reactivity, Pulmonary Inflammation, and Chemokine Expression". *Toxicological Sciences* 72.2, pp. 339–346.

Landrigan, Philip J., Richard Fuller, Nereus J. R. Acosta, Olusoji Adeyi, Robert Arnold, Niladri (Nil) Basu, Abdoulaye Bibi Baldé, Roberto Bertollini, Stephan Bose-O'Reilly, Jo Ivey Boufford, Patrick N. Breysse, Thomas Chiles, Chulabhorn Mahidol, Awa M. Coll-Seck, Maureen L. Cropper, Julius Fobil, Valentin Fuster, Michael Greenstone, Andy Haines, David Hanrahan, David Hunter, Mukesh Khare, Alan Krupnick, Bruce Lanphear, Bindu Lohani, Keith Martin, Karen V. Mathiasen, Maureen A. McTeer, Christopher J. L. Murray, Johanita D. Ndahimananjara, Frederica Perera, Janez Potocnik, Alexander S. Preker, Jairam Ramesh, Johan Rockström, Carlos Salinas, Leona D. Samson, Karti Sandilya, Peter D. Sly, Kirk R. Smith, Achim Steiner, Richard B. Stewart, William A. Suk, Onno C. P. van Schayck, Gautam N. Yadama, Kandeh Yumkella, and Ma Zhong (2018). "The Lancet Commission on Pollution and Health". *The Lancet* 391.10119, pp. 462–512. DOI: 10.1016/S0140-6736(17)32345-0.

Lauer, Stephen A, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler (2020). "The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application". *Annals of Internal Medicine* 172.9, pp. 577–582.

Le Monde (2020). "Il serait inacceptable de sortir demain de la crise du Covid-19 pour mourir de la pollution de l'air".

Lyapustin, Alexei and Yujie Wang (2018). "MCD19A2 MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G Global 1km SIN Grid V006". DOI: 10.5067/MODIS/MCD19A2.006.

Ogen, Yaron (2020). "Assessing Nitrogen Dioxide (NO2) Levels as a Contributing Factor to Coronavirus (COVID-19) Fatality". *Science of The Total Environment* 726, p. 138605. DOI: 10.1016/j.scitotenv.2020.138605.

Peng, Lu, Xiuge Zhao, Yan Tao, Shengquan Mi, Ju Huang, and Qinkai Zhang (2020). "The Effects of Air Pollution and Meteorological Factors on Measles Cases in Lanzhou, China". *Environmental Science and Pollution Research*, pp. 1–10.

Persico, Claudia L and Kathryn R Johnson (2020). "Deregulation in a Time of Pandemic: Does Pollution Increase Coronavirus Cases or Deaths?" *Journal of Environmental Economics and Management (forthcoming)*.

Rivellese, Felice and Edoardo Prediletto (2020). "ACE2 at the Centre of COVID-19 from Paucisymptomatic Infections to Severe Pneumonia". *Autoimmunity reviews* 19.6, p. 102536.

Robert Koch Institut, Bundesamt für Kartographie und Geodäsie (2020). *CSV Mit Den Aktuellen Covid-19 Infektionen pro Tag (Zeitreihe)*. URL: https://www.arcgis.com/home/item.html?id=f10774f1c63e40168479a1feb6c7ca74 (visited on 06/09/2020).

Sager, Lutz (2019). "Estimating the Effect of Air Pollution on Road Safety Using Atmospheric Temperature Inversions". *Journal of Environmental Economics and Management* 98, p. 102250. DOI: 10.1016/j.jeem.2019.102250.

Son, Ji-Young, Kelvin C. Fong, Seulkee Heo, Honghyok Kim, Chris C. Lim, and Michelle L. Bell (2020-11-20, 2020). "Reductions in Mortality Resulting from Reduced Air Pollution Levels Due to COVID-19 Mitigation Measures". *Science of The Total Environment* 744, p. 141012.

The New York Times (2020). *Coronavirus (Covid-19) Data in the United States*. URL: `https://github.com/nytimes/covid-19-data` (visited on 06/09/2020).

Travaglio, Marco, Yizhou Yu, Rebeka Popovic, Liza Selley, Nuno Santos Leal, and L. Miguel Martins (2021). "Links between Air Pollution and COVID-19 in England". *Environmental Pollution* 268, p. 115859.

UK Government (2020). *Coronavirus (COVID-19) in the UK*. URL: `https://coronavirus.data.gov.uk/` (visited on 06/09/2020).

Wooldridge, Jeffrey M (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Wu, Xiao, Rachel C. Nethery, Benjamin M. Sabath, Danielle Braun, and Francesca Dominici (2020). "Exposure to Air Pollution and COVID-19 Mortality in the United States: A Nationwide Cross-Sectional Study". *medRxiv*, p. 2020.04.05. 20054502. DOI: `10.1101/2020.04.05.20054502`.

Yang, Jing, Ya Zheng, Xi Gou, Ke Pu, Zhaofeng Chen, Qinghong Guo, Rui Ji, Haojia Wang, Yuping Wang, and Yongning Zhou (2020). "Prevalence of Comorbidities and Its Effects in Patients Infected with SARS-CoV-2: A Systematic Review and Meta-Analysis". *International Journal of Infectious Diseases* 94, pp. 91–95.

Ye, Qing, Jun-fen Fu, Jian-hua Mao, and Shi-qiang Shang (2016). "Haze Is a Risk Factor Contributing to the Rapid Spread of Respiratory Syncytial Virus in Children". *Environmental Science and Pollution Research* 23.20, pp. 20178–20185.

Zhou, Fei, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. (2020). "Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China: A Retrospective Cohort Study". *The Lancet*.

Zhu, Yongjian, Jingui Xie, Fengming Huang, and Liqing Cao (2020). "Association between Short-Term Exposure to Air Pollution and COVID-19 Infection: Evidence from China". *Science of The Total Environment* 727, p. 138704. DOI: `10.1016/j.scitotenv.2020.138704`.

Chapter 7

Synthesis and outlook

This dissertation is concerned with the identification of economic parameters using counterfactual causal inference. Building on recent research on the caveats of fixed effects regressions, I demonstrate how to analyze standard panel econometric research settings in terms of counterfactuals, reconstructing potential outcomes for cohorts of treated units, based on suitable controls. This approach allows for transparency of the parallel trends and stable unit treatment value assumptions, and a structured discussion of the exogeneity of treatment itself or its timing. I apply this to recover estimates of the causal effects in a range of settings relevant for sectoral environmental and climate policy. In this Synthesis Chapter 7, I revisit the assumptions of an unbiased generalized difference-in-differences design and discuss their role in some further empirical settings in environmental economics (Section 7.1). I provide a more detailed discussion of the *generalizability* of counterfactuals, based on my own research and additional examples from environmental economics, such as the effects of Pigouvian taxes.

Why are we interested in causal effects? We may, for instance, want to know if a particular intervention worked. For this, we need an unbiased counterfactual. However, we also want to know if a policy would work in the future, in a different context, and how large the effect would be at lower or higher treatment intensities. This is why the *external validity* of our comparisons between treatment group and counterfactual needs to be examined. When David Hume wrote down his definitions of causality that I discussed in Chapter 1, he wanted to propose a framework that allows us to induce some generalizations from observations. In his second attempt at defining causality, in which he also proceeds to introduce the notion of a counterfactual, he first proposes: "We may define a cause to be an object followed by another, and where all the objects similar to the first are followed by objects similar to the second" (Hume 1748). In a nutshell, Hume is saying that we can make predictions about other contexts based on an observed correlation. This is also the principle behind many machine learning algorithms: if changes in a variable consistently follow a combination of changes in other variables, we can make confident predictions out of samples, i.e., for the future or for other populations. However, this prediction approach only works well under a suitable combination of two elements: (i) there needs to be enough data to fit a function to the data that is representative of the latent mechanisms driving an outcome,[1] (ii) the latent factors driving the prediction do not change from the training sample to the prediction sample. If we find a truly unbiased counterfactual using the methods discussed in this dissertation, we do not have to worry about the latent factors. We have uncovered a causal relationship which is free of any latent, unobserved variable bias. However, we still cannot be sure that the relationship would be meaningful in other

[1]For a synthetic control, that means a sufficient number of pre-treatment time steps used for matching is needed in order to make sure that a good pre-treatment fit is not due to chance.

contexts, because our sample may not be representative for these other populations (in the future or in a different place) or because the treatment may look different there (in its intensity or in its implementation). This is the issue of generalizability, which I will discuss in the next section.

7.1 Common insights

Generalizability

Kahneman (2011) famously noted that "when faced with a difficult question, we often answer an easier one instead, usually without noticing the substitution". Researchers often have to pick a question for which they have data and for which an unbiased counterfactual can be constructed. In Chapter 1 (Introduction), I have discussed how to use contemporaneously untreated units as controls to build a counterfactual in a setting of staggered and repeated treatment. This discussion was focused on avoiding biased estimates, particularly due to violations of the parallel trends assumption. However, finding an unbiased counterfactual can come at the expense of its external validity. This means that the estimates are not representative for other contexts (either in the cross-section or in time), or even for the broader research question which we fundamentally want to answer. In Section 1.2, we discussed this problem in the context of the empirics of carbon pricing instruments. It is difficult to extrapolate from empirical findings on emission reductions based on existing carbon pricing schemes, because they analyze data from limited temporal and cross-sectional contexts. Since, in many interventions, carbon prices were small, it is not surprising that their total effect when investigating treatment as a binary indicator variable is small.[2] Some studies estimate dose-response relationships; for instance, the price elasticity of fossil fuel demand. However, here researchers encounter a clear limitation of studies built on the potential-outcomes approach, the Rubin causal model discussed in Chapter 1, which consists of a comparison of the means of a treatment group and its counterfactual (Equation 1.1). Researchers implement this comparison of means in regression-form estimating linear models, such as Equation 1.2. Again, we cannot confidently extrapolate from these studies, since we do not know if the linear functional form holds out of sample, i.e., for levels of pricing that we do not observe empirically. Researchers can, of course, estimate non-linear models. This should, however, be grounded in ex-ante theoretical reasoning, rather than ad-hoc decisions on functional form, which increase the risk of "overfitting". We discuss these external validity issues in this section.

[2]This is, however, how these studies may be perceived in public discourse. Communicators and readers may take a shortcut and conclude "carbon pricing does not work".

Choosing a control group Empirical evaluations of interventions often report results with sentences of the type: "Treatment D increases Y by X%". In this dissertation, I have argued that regression-based research designs implicitly construct a counterfactual that can be unbiased under certain conditions. However, the choice of the control group also determines if the findings are applicable to other contexts and tell us something about the effect in which we are fundamentally interested. Therefore, an important guiding question for a critical read of a research paper is: "Compared to what?".

Consider again the case of the hypothetical tax on deforestation in certain Brazilian municipalities that is rolled out in a staggered manner.

- Should we compare only municipalities that eventually introduce the tax?
- Should we compare them to all Brazilian municipalities?
- Should we compare municipalities that have different deforestation tax levels?
- Should we find control municipalities that match the observable characteristics of treated ones?
- Should we use control municipalities from other Latin American countries to keep some of the general equilibrium effects at the national level?
- Should we try to investigate national-level cases of deforestation taxes (or similar) from several regions and decades?
- Should we dispense with control units entirely and just look for trend breaks in the time series?[3]

Beyond the issue of bias, we see that these different counterfactuals are also answering different questions about the effectiveness of a deforestation tax and that there can be a trade-off between bias and external validity. Some counterfactuals can be better suited to teach us what would happen – if a deforestation tax was introduced at a different level, at a different place, and at a different time. They may, however, create more bias because they are not similar enough to the treated units. However, if they are too similar, we tend to have spillovers that lead to bias, and autocorrelation that makes estimates seem more precise than they are (see Section 1.2).

Researchers should therefore provide a detailed discussion of the counterfactual constructed in a study, the question it answers, and the sensitivity of results to changes in the counterfactual. The increased availability of micro-data at a fine resolution helps to improve identification, but it may lead to the neglect of other questions for which there is worse data or a less clean counterfactual.[4]

[3]In finance, the term *event study*, which is now often used to describe a difference-in-differences design with relative time indicators, has typically referred to time series studies based on trend breaks.

[4]The potential trade-offs between internal and external validity in economic research have been discussed extensively in the context of RCTs.

Some observers have described this as a version of the lamp post problem, in which a researcher examines a narrow context because it provides an unbiased counterfactual, for instance due to a natural experiment, and misses some broader, potentially more important questions, just like the drunk person in the parable searching for their key under the streetlamp, despite the fact that they know they dropped it somewhere else. One example is the literature on the *natural resource curse*. Early attempts to investigate the reasons for many countries being poor despite their substantial natural resources were largely based on cross-country regressions (van der Ploeg 2011). More recent literature, enabled by new data sources, examines the question based on within-country and micro-level variation (Cust and Poelhekke 2015), with some studies aiming to provide microfoundations to specific channels. For instance, Cappelen et al. (2021) showed in a field experiment in Tanzania that knowledge about the discovery of natural gas makes people expect more corruption in the future, but does not make them less trusting or more corrupt in the present. While some of the mechanisms investigated at the local level, such as local labor markets, institutions, and government spending, can help improve our understanding of the broader economics of the resource curse, many of the proposed channels of a resource curse should predominantly be detectable at the national level. Some mechanisms, such as the exchange rate and the financial system, are features of the national economy and will, therefore, not be captured by within-country studies. Other mechanisms, such as conflict and corruption, can have local origins, but will, if they are relevant enough to cause a resource curse, have repercussions on the national level and, therefore, "get controlled away" when only within-country variation is used.[5] Consider the case of the opening of new coal mines across a country. This could lead the places around coal mines to increase their GDP compared to other, similar places without a new coal mine. As these control places could be experiencing negative spillovers because labor or capital move to treated places, researchers could be measuring an effect, even if on a net basis there is no improvement at the national level. However, even if spillovers are accounted for, if the additional production of coal leads to an appreciation of the exchange rate, the national economy might be on a net basis worse off than in the counterfactual without the new coal mines.

In Chapter 2, I investigate the effects of the Indonesian palm oil boom on structural change in Indonesian districts. I use new investments in palm oil processing facilities (mills) to identify spillovers on unrelated manufacturing plants. Since palm oil is a labor-intensive sector, one could expect a crowding-out of manufacturing. I show that the opposite is the case: despite some increases in blue-collar wages, manufacturing plants exposed to the shock increase their sales and productivity. As I find substantial labor migration, I also demonstrate that these results are robust to excluding neighbour district from the control group for each cohort. However, additional spillovers between treated and control units might be present in general equilibrium and at a higher level of geographic aggregation, such as the regional or national level. For instance, local palm oil booms could pull national political attention or financial capital allocated by regional banks

[5]This could also be considered an issue of "bad control" (Angrist and Pischke 2008) or spillovers.

away from other sectors. If the productivity of these sectors grows faster, for instance because of learning-by-doing, or if these sectors have positive externalities on the rest of the economy, dynamically, this would decrease overall economic activity compared to a national counterfactual. Therefore, even though I have found a suitable within-country control group to detect positive local spillovers, it is not clear what the results mean for overall, regional and national structural change in Indonesia.

Related literature in development economics and economic history that investigates the long-run effects of colonial regimes raises similar questions. While this literature documents the negative effect of extractive colonialism on institutions (Acemoglu et al. 2001), some authors have also found that those areas that have seen more colonial investments are also more developed today (Dell and Olken 2020; Jedwab and Moradi 2015). Some observers may conclude from these latter results that colonialism was good for economic development. However, this is not what they show. They mainly document that there is a long path-dependence in the geographic distribution of economic activity. This path dependency is confirmed for colonial cash crop plantations (Roessler et al. 2020) that make plantation areas better off compared to a within-country counterfactual, notably also crowding out development in neighboring areas. However, additionally, and potentially more importantly, with these types of colonial interventions, there may be institutional and macro-economic forces acting at the national level, that within-country studies do not capture. These national-level phenomena may well be stronger than local effects, but they are harder to identify econometrically. Broadly, there could be two solutions to this problem: (i) using higher-resolution data to estimate the micro parameters of a more structural model of the aggregate economy, or (ii) using a unit of observation, i.e., regions or countries, that captures general equilibrium effects.[6] The second approach essentially consists in running cross-country regressions, which is the methodology used in the first phase of the resource curse literature.

Fortunately, new approaches, such as the *synthetic control method*, which has been called "arguably [the] most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens 2017), now allow researchers to address some of the threats to identification that have concerned critiques of cross-country growth regressions, as they allow the construction of credible counterfactuals, even for entire countries. Intuitively, the method builds a synthetic control unit for each treated unit, based on a mix of control units from a donor pool. In Section 7.2, I explain the use of this method in policy evaluation, particularly if the synthetic counterfactual is specified before the policy is implemented. I use the example of an evaluation of the German carbon price (*Brennstoffemissionshandelsgesetz*), for which my co-authors and I pre-registered a pre-analysis plan in December 2020 (Naumann et al. 2020), thus defining the construction of the counterfactual before seeing outcome data from periods after treatment.

[6]Note that this discussion focuses on the common research setting that consists in using sub-national variation to investigate national phenomena. However, treatments can also vary by economic sector or demography, rather than only geographically.

Choosing a treatment definition To capture unbiased estimates for an economic parameter researchers seek out research contexts that allow to construct unbiased counterfactuals. Often, this also means considering different treatment definitions.

In her canonical study on the economic returns to a large-scale school construction program in Indonesia, Duflo (2001) presents both a simple difference-in-differences specification with a binary treatment indicator and a generalized difference-in-differences analysis recovering a dose-response relationship. The first type of analysis, using the binary treatment indicator, has the advantage of being able to show a simple line chart comparing the mean of treated group and control group over time. The second type uses the additional variation from treatment intensity, thereby increasing statistical power and therefore, for instance, also helping to avoid Type II error (false negatives). Ideally, researchers show and discuss both approaches.

I use both approaches (binary and continuous treatment) in Chapter 3 about the effects of provisional bike lanes on cycling traffic. These pop-up bike lanes were rolled out in a range of European cities in 2020. I implement a generalized difference-in-differences design using variation in both treatment intensity and treatment timing to estimate the infrastructure-elasticity of cycling: By how much does cycling increase for every kilometer of additional infrastructure? I also estimate the change in cycling for an average pop-up bike lane program using a dummy for treatment. Note that, while the kilometer-based analysis at first glance seems to give us better external validity to judge similar future programs in other contexts, we may also consider the dummy treatment the more prudent specification. This is because the bike lane treatment is not only based on the kilometers of infrastructure themselves, but also on the overall local public attention to the program. However, at the city-level, the treatment intensity still holds valuable information that can be exploited, since media attention will be lower when a program is smaller.

Another concern is that it is not only the choice of an identification strategy, such as the levels of fixed effects, that influences the counterfactual, but also the type of treatment variation used. Consider the empirical literature on the effects of climate change. Researchers have used panel data and included time-varying fixed effects in their regressions to remove potential omitted variable bias.[7] Others have relied on simple cross-sections or long-differences. A long-difference design exploits changes between two time steps with more distance than in a typical panel model that builds on short-term variation. The central, ongoing debate in this empirical literature is concerned with the question of how to define the climate treatment to ensure it generates an estimate that is both unbiased and has external validity for the impacts of future climate change (see Auffhammer (2018) for an overview).

The first type of specification, the panel approach, will capture short-run variations in climate variables, i.e., weather shocks, while the second type of specification will capture

[7]For a critical discussion of arguments for panel approaches based on omitted variable bias in the climate impacts literature, see my discussion in Section 7.2.

more long-run climate change effects, including adaptation. Clearly, the higher the resolution of fixed effects is in these models, the more they remove lagged effects from previous changes in climate and previous weather shocks that could have influenced a wide range of economic and social variables, and thereby also selection into treatment. If researchers are interested in high statistical power and a clean weather shock, not influenced by lagged effects of climate, it is therefore sensible to estimate panels with high temporal and geographical resolution. Critics of this approach have noted that that the simple panel regression approach will "control away" all effects of adaptation. They also do not allow for comparisons between places with large temperature differentials, since country or regional fixed effects make sure that comparisons are only made within a restricted geographical area.[8] In contrast, the second type of specification, the cross-sectional approach, captures differences in climate and the lagged effects of these differences[9] that may not generate results with good external validity. Researchers have proposed regression specifications that capture elements of both climate and weather (Kolstad and Moore 2020). Starting with McIntosh and Schlenker (2006), researchers have used non-linear specifications of weather to capture a combination of within-unit and cross-sectional variation. Similar in spirit to the deconstruction of the two-way fixed effect model that we discuss in Section 1.2, Mérel and Gammans (2021), we decompose this non-linear climate impacts panel regression and show that they generate a weighted average of short- and long-run effects. Depending on the structure of the panel and the weather variation used after conditioning on controls and fixed effects, i.e., after partialing them out from both the outcome *and* the weather treatment, the approach yields a different mix of weather and climate effects. In other words, the regression generates a range of comparisons between shocked units and their counterfactuals, which are constructed from other units and within units. When researchers investigate different identification strategies, they often also implicitly alter their treatment definition. Decomposing panel regressions and comparing results from different specifications of a treatment are both helpful tools to create more transparency about what kind of elasticity or parameter is estimated.

Revisiting difference-in-differences assumptions

Parallel trends

Difference-in-differences designs compare treated to control units over time. They provide unbiased results, if it is possible to model and remove a common trend between treatment and control group based on the trend of the control group (see Equation 1.2), i.e., the control group has to be a good counterfactual for the treatment group. This is the parallel trends assumption that we discussed in Section 1.2. If the parallel trends assumption does

[8]Note that these fixed effects also remove some more aggregate, macro effects, such as national general equilibrium effects, from the variation.

[9]If only cross-sectional variation is used, that intuitively includes the whole history of the Earth. Note that, in this approach, we also cannot check if the correlation between climate and economic outcomes is due to chance, because we cannot run the whole history of the Earth as an experiment again.

not hold, estimates will be biased. For instance, if two variables are trending differently, with no changes in the trend around treatment, with sufficient statistical power, we will detect an effect, even if there is none. We have shown in Section 1.2 that fixed effects panel regressions are a generalized version of difference-in-differences. For these regressions to provide unbiased estimates, they also need to satisfy some form of generalized parallel trends assumption. I have shown how to disentangle this setup into a stacked difference-in-differences design that allows investigation of the parallel trends assumption, both visually and in regression-form. Chapter 2 critically builds on the fact that manufacturing plants exposed to a new palm oil mill are on a parallel trend to similar manufacturing plants. Note that this parallel trend is not a result of a matching procedure, but holds conditional on a set of standard fixed effects from the literature that works with firm micro-data. In the Chapter, we also discuss the issue of a slight Ashenfelter dip[10] that appears when using a stricter set of fixed effects and provides estimates, excluding those years for which estimates are already subject to anticipation effects.

A visual demonstration of parallel pre-trends is now a pre-requisite for simple difference-in-differences designs. It could become so for generalized difference-in-differences, too, therefore applying to many types of more complex panel regressions. However, parallel pre-trends are only at most half the battle of a credible identification strategy. Pre-trends that are not parallel are a clear indication of some kind of endogeneity, but pre-trends that are parallel are not sufficient to rule out endogeneity. This is because there can be omitted variable bias or reverse causality that only becomes salient in the outcome after treatment. Therefore, in order to make the parallel trends assumption more credible, researchers can argue that treatment itself or treatment timing are exogenous. I will now discuss this issue and how we approach it in the main Chapters of this dissertation.

Exogenous treatment (timing)

The difference-in-differences design can be unbiased, even if treatment and control group are different at baseline. However, if researchers cannot credibly argue that treatment is exogenous conditional on covariates, this also makes it more difficult to build an argument that the parallel trends assumption holds. Remember that the parallel trends assumption is that treatment and control group would have evolved on a parallel trend in the absence of treatment. We do not know the treated unit in its counterfactual untreated state after treatment has happened. Therefore, even if we can show parallel pre-trends, we have to make a convincing argument that this would have continued, if treatment had not happened. Clearly, the parallel trends assumption would, for instance, not hold, if treatment were endogenous, but this endogeneity only appears in the outcome after treatment. Importantly, reverse causality can have this feature, when a treatment reacts contemporaneously to a shift in the outcome or if treatment anticipates a shift

[10] An Ashenfelter dip (1978) refers to a reduction in an outcome just before treatment, that can indicate changes in behavior in anticipation of treatment, which can be a source of bias (1978).

in the outcome.[11] In the staggered adoption case, an alternative to an identification argument based on parallel pre-trends is to assume that treatment adoption dates are randomly assigned (Athey and Imbens 2018).[12] In Chapter 3, we do both: we show that provisional cycling infrastructure increases cycling rapidly and substantially. We build our identification strategy on two strands of arguments. First, we make a standard difference-in-differences comparison between treated and control cities over the months preceding and following the treatment hard-coded to March 2020. This means that we do not use variation in treatment timing for identification but fully rely on untreated control cities as a suitable counterfactual. Second, we also make comparisons within a narrower window of time around the day of treatment, using daily variation in cycling and different treatment intensities in a generalized difference-in-differences setting.

However, when high-frequency data is not available, endogeneities might stay hidden in aggregate time-series. Then researchers have to rely heavily on arguments around the exogeneity of treatment for all individual treatment cohorts, conditional on fixed effects and controls or on the exogeneity of treatment timing. The stacked difference-in-differences design can help to make this assumption more plausible. It makes both the shocks and the counterfactuals constructed more transparent and allows for a structured discussion of the relationships between the two. The stacked design decomposes a staggered adoption setting into different sub-analyses for each cohort, or even for each treated unit. It allows additional fixed effects to be included that remove time trends and time-invariant differences between treatment and control group for each sub-analysis. If there is enough statistical power, researchers can also investigate treatment dynamics around each of the individual cohorts and check the robustness of results when those cohorts for which pre-trends are not parallel are excluded.

In Chapter 2, in addition to showing parallel pre-trends, we make an identification argument based on the exogeneity of investment decisions by large palm oil groups to local conditions that could make treatment timing endogenous. We also run the analysis for a sample with only the largest corporations, thereby excluding those investments that could come from smaller, more local actors potentially more responsive to local economic conditions.

[11] Note that this is another problem with the Humean conception of causality that is also present in methods building on *Granger causality* (Granger 1969): two factors can be causally related, even if a measurable change in the effect does not chronologically follow a measurable change in the cause. For instance, economic agents can expect a treatment and create shifts in the outcome in anticipation. In turn, lagged explanatory variables sometimes used as instruments for a treatment should not be assumed to be exogenous because they can have direct effects on an outcome or effects through a different channel than the actual treatment, thereby violating the exclusion restriction (Bellemare et al. 2017).

[12] Athey and Imbens (2018) underline "Commonly made common trend assumptions [. . .] follow from some of our assumptions, but are not the starting point".

SUTVA

In the presence of spillovers between treated and control units, the stable unit treatment value assumption (SUTVA) is violated. Intuitively, this happens because spillovers "contaminate" the control group, which, depending on the direction of the spillover effect, can lead to under- or overestimation of the treatment effect. Consider the case of a study on the effectiveness of a vaccine, comparing vaccinated people with a control group. Bias away from the true effect would occur in the case of a vaccine that provides herd immunity to a community that includes control units. Bias in the direction of the true effect would occur if individuals communicating that they have been vaccinated led to riskier behaviour in their community, leading to more infections in the control group and to researchers overestimating the effectiveness of the vaccine.[13]

Many of the phenomena studied in environmental economics are public goods, such as air or water quality. Therefore, externalities and spillovers are central to the empirics of the field (Deschenes and Meng 2018). However, when spillovers cannot be directly measured or modeled, this leads to biased effects. Similar problems arise when studying highly integrated markets. For instance, if markets have substantial mobility and low friction or homogeneous goods, spillovers across study units and general equilibrium effects at more aggregate levels of the economy will be a concern. Deforestation is again an intuitive example for this issue. Many of the key commodities that drive deforestation (vegetable oil, protein feed) are homogeneous goods and in some cases virtually perfect substitutes for each other. Their production function is comparatively simple, they benefit from agglomeration effects only to a limited extent, and in many countries vast areas are theoretically suitable for their cultivation. Therefore, we should expect a local treatment like a tax on deforestation to have spillovers, often even beyond national borders through global commodity markets. This makes it more complicated to find a suitable counterfactual. The literature that investigates what drives and what stops deforestation has paid particular attention to the issue of leakage (Meyfroidt et al. 2020).

In our study on the effects of palm oil investments on the manufacturing sector (Chapter 2), we also find substantial in-migration into the district. Therefore, we investigate spillovers between neighboring districts and run the main analysis removing neighboring districts of those treated in each cohort. The stacked difference-in-differences design allows "contaminated" control units to be mechanically removed from the sample for each cohort. These robustness checks confirm our main result. In the stacked design, researchers can also build a treatment group consisting of units expected to be exposed to spillovers and "uncontaminated" control units. In Chapter 6 on the effects of air pollution on Covid-19 outcomes, we examine the role of spillovers by running a robustness check with a sample

[13]Bias in the direction of the true effect would also occur if a vaccine only protects the vaccinated person and does not break transmission. If the vaccinated person engages in riskier behavior because of their increased protection, they might be driving more community transmission, thereby literally contaminating their community and therefore also control units. This is, of course, not the case if treatment is successfully blinded to the participants.

restricted to districts with strict stay-at-home orders. This removes infection spillovers from people travelling across district borders.

7.2 Outlook: new research practices

This dissertation has focused on constructing credible counterfactuals. This has also been the mission of the "credibility revolution" in economics. However, the replication crisis in psychology has alerted economists to another problem that might even be obscured by ever more elaborate empirical strategies: What if many of our new well-identified papers do not replicate? Researchers have started investigating this question using prediction markets (Gordon et al. 2020). They ask participants to gauge the likelihood that scientific papers replicate and run replications for a sub-sample to verify the predictions. Prediction market participants consider economics the social science whose papers are most likely to replicate (Gordon et al. 2020). Still, from a sample of economics papers from top journals, they predict that only 2/3 of papers would hold up to replication.[14] Before seeing any papers, they expect that average prediction rates rise over time from 43% in 2009/2010 to 55% in 2017/2018 (Gordon et al. 2020). After seeing the papers, they do not think that there was a substantial improvement over time anymore. However, they also predict a slightly higher level of replicability than in 2009 (53.4% in 2009 and 55.8% in 2018) (Menard 2020). Why do papers not replicate? It may be because researchers made coding errors or because they cherry-picked cleaning procedures or regression specifications.[15] The problem is exacerbated because of *publication bias*, the fact that many journals tend to prefer findings that are statistically significant at a given p-value level. A problem with replication could even arise in a study context where the true population effect is small and estimates in a study are extreme due to chance. Another reason for a failure to replicate could be related to external validity; for instance, if a new study population is too different from the original study population or if a study population changes so strongly between two studies that results do not replicate anymore.

In this section, I discuss future avenues for research practices that could help improve replication in economics, with a particular focus on pre-analysis plans, which can help counter both the issues of specification search and publication bias.

Similar to predictions made by machine learning algorithms, even if studies replicate, it does not mean that their result can be interpreted as causal. A successful replication may just prove that the initial study picked up a signal, rather than noise; the problem of bias

[14]Both numbers are based on a prediction market, in which participants bet prize money on the replication of papers they read. The results of this study are not published yet. However, such prediction markets have been shown to be rather accurate (Camerer et al. 2018; Dreber et al. 2015). For participants' expectations about different disciplines before seeing the papers to be judged, see Figure 1 in Gordon et al. (2020). For the predictions after seeing the papers, see Figure 13 in Menard (2020).

[15]Simmons et al. (2011) call this "research degrees of freedom" and Gelman and Loken (2017) the "garden of forking paths". If researchers cherry-pick specifications to obtain a statistically significant result for a given p-value level, this is also called "p-hacking".

may remain. In this dissertation, I have focused on producing unbiased counterfactuals. The "credibility revolution" has shifted applied economics' focus from the extensive use of control variables in regressions to more parsimonious models restricted to a source of exogenous variation and some fixed effects. However, there are contexts with many potential confounding factors or some highly salient omitted factors that vary at a high resolution below any fixed effects levels. In these cases, it is prudent to include control variables in regressions. The difference-in-differences in regression form, that we discussed in Chapter 1, Section 1.2 allows for the inclusion of additional covariates $X_i t$. However, in practice, this means partialing out variation in $X_i t$ from both the treatment and the outcome. This makes it complicated to express the regression result as a weighted average of comparisons in group means representing treated (treatment group) and counterfactual (control group) potential outcomes (Equation 1.2 in Goodman-Bacon (2018)), i.e., the empirical setting cannot easily be interpreted with the Rubin Causal Model. In this Section, we therefore also discuss an alternative causal framework based on *directed acyclic graphs* (Pearl 2009).

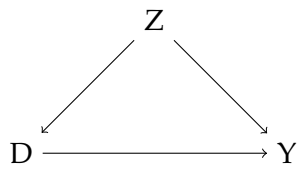
Choosing controls with directed acyclic graphs

Since the "crediblity revolution", randomized experiments have been the benchmark econometric studies. Identification strategies are typically built on plausibly exogenous variation in a treatment. Whenever exogenous treatment is not a plausible assumption, researchers also try to express their results as a comparison between a treatment and a control group. In this dissertation, I advocated the latter approach, i.e., that researchers should build their research designs on unbiased and relevant counterfactuals. However, I have also discussed that, even when researchers find parallel pre-treatment trends between treatment and control group, they still have to make an argument that there are no endogeneities that become salient in the outcome only after the time of treatment. Researchers can for instance do this by discussing potential sources of endogeneity in treatment timing. They should also be able to argue why any visible level differences between treatment and control group will not be the driver of a divergence in trends after treatment (McKenzie 2020).

Directed acyclic graphs (DAGs) provide a framework that helps to think about the causal relationships between outcome, treatment and third variables in a structured way. DAGs are a visible representation of the relationships between variables in an empirical setting, that enable a structured discussion about the type of bias that can be created by including or leaving out individual variables (see Figures 7.1, 7.2, and 7.3). While, from the quasi-experimental point of view, a researcher has to tell a convincing story as to why no potential determinant of the outcome can be influencing treatment, DAGs allow consideration of all relevant variables in a setting and assess the type of biases they can introduce. This may be omitted variable bias from a third variable affecting both treatment and outcome (in DAGs this is called a "backdoor"). However, third variables can also be mediators;

introduce post-treatment bias; may be so-called "colliders", which I discuss below; or may be mediators in reverse causality. DAGs also facilitate a structured argument about the direction and the magnitude of bias introduced by measurement error (Hernán and Cole 2009). In this dissertation, I have therefore used these graphs in the design phase of my research. While it is rare to see them in the bodies, or even appendices, of economics papers,[16] they may see increased adoption by economists in the near future.

Figure 7.1: Omitted variable bias from a backdoor in a directed acyclic graph (DAG)



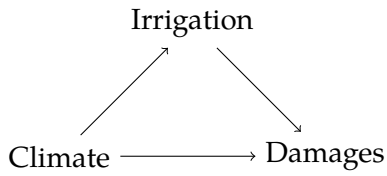
Notes. This figure illustrates omitted variable bias from a "backdoor" path in a directed acyclic graph (DAG). D is a treatment or exposure, Y is an outcome, and Z is a third variable that affects both D and Y, thereby creating omitted variable bias. In the DAG vocabulary, Z opens a backdoor between D and Y (Pearl 2009). This figure is created with the tool *DAGitty* (Textor et al. 2016).

DAGs may seem trivial, but they have important complementarities with the potential outcomes framework. The main difference is that they allow researchers to put the direction of causality at the center of their analysis of endogeneity. Thus, DAGs help researchers to be more specific about omitted variable bias. For instance, in a regression of outcome Y on treatment D, omitted variable bias can be created by a "backdoor", i.e., when a third factor Z causes both Y and D. In DAG, this is illustrated by an arrow pointing from Z to both Y and D (Figure 7.1). I illustrate the usefulness of DAGs in the context of the estimation of climate damage functions. In the climate impacts literature, omitted variable bias is often defined as a correlation between the error term of a regression and the climate treatment. One of the main criticisms of the early cross-sectional approach by Mendelsohn et al. (1994) was that it was likely to suffer from omitted variable bias; for instance, by leaving out variables such as irrigation (Auffhammer 2018). However, if we draw a DAG for this example (Figure 7.2), we see that irrigation cannot cause omitted variable bias through a backdoor, since it is an effect of climate rather than a cause. Therefore, in this literature, when researchers talk about omitted variable bias, they must mean something different. Omitted variable bias is present when an estimated parameter deviates from the true population parameter in which we are interested (Angrist and Pischke 2008). Thus, if researchers claim that there is omitted variable bias from irrigation in this setting, that means that they are only interested in the effect of temperature or precipitation variation on an outcome after removing causal channels that run through irrigation. The DAG shows that, depending on how we define our treatment effect, there may or may not be bias from this form of adaptation.

Does this mean that all papers in this literature after Mendelsohn et al. (1994) that capture a full long-run climate effect have been futile? On the contrary: First, the literature has provided us with a wealth of new knowledge about the magnitude and mechanism

[16]Epidemiology is a field that has been faster to incorporate DAGs (Hernán and Robins 2020).

Figure 7.2: Irrigation is a mediator between climate and damages



Notes. This figure illustrates the causal relationships between the treatment $Climate$, the outcome $Damages$ and the third variable $Irrigation$ in a directed acyclic graph (DAG). There is no arrow from $Irrigation$ to $Climate$, since a causality in this direction is not plausible. $Irrigiation$ is a mediator between $Climate$ and $Damages$. This figure is created with the tool *DAGitty* (Textor et al. 2016).

of human reactions to different weather phenomena, even if some of them include the *extensive margin*, i.e., adaptation, only to a limited extent, which on the plus side also allows us to assess the extent to which human populations have been able to adapt to past changes in the climate. Second, since projecting climate damages in the future is fundamentally a problem of external validity, it is valuable to get estimates based on a wide variety of counterfactuals to facilitate a scholarly discussion. Some researchers now also provide a range of estimates from different estimation approaches (Kalkuhl and Wenz 2020).

DAGs also illustrate that more control variables do not always mean better identification. Economists now typically refrain from giving control variables in their regressions causal interpretations,[17] or any interpretations at all. Many specifications in economics papers, however, still include vectors of control variables.[18] Many researchers also know that controlling on a variable that is itself an effect of the outcome is a case of "bad control" (Angrist and Pischke 2008) and that a variable that lies between treatment and outcome, i.e., that is an effect of the treatment and a cause of the outcome, is a mediator. Controlling for such variables will change the estimates for the effect of the treatment in an undesired way. There is a third type of problem introduced by controlling too much, which is less known: conditioning on a collider.

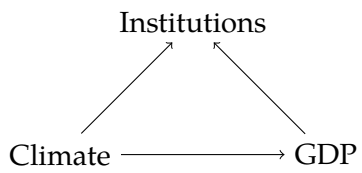
The most simple example of a collider is a third variable Z that is caused by both the outcome and the treatment.[19] If researchers condition on such a collider, they effectively look at a relationship between outcome and treatment at different levels of Z. This relationship may be different and even have a different sign from the relationship in the overall population without conditioning on Z. Figure 7.3 shows an example for such a collider in climate impact research: institutions. Since institutions are at least partially driven by both climate and GDP, researchers introduce collider bias into their estimates, if

[17]This is because control variables can be subject to endogeneities that can potentially even change the sign of their coefficient (Hünermund 2019).

[18]Note that including controls can increase the precision of estimates. The inclusion of a set of control variables at the ad-hoc discretion of the researcher, however, also expands the researcher degrees of freedoms discussed in Section 7.2.

[19]Even less intuitively, collider bias can also be present when the collider is not directly caused by treatment and outcome, but by variables that are themselves causes of treatment and outcome. This is called *M-Bias* (Elwert and Winship 2014).

Figure 7.3: Institutions are a collider variable for the relationship between climate and GDP



Notes. This figure illustrates a simple, hypothetical example of collider bias in the context of the climate impacts literature. $Institutions$ are both at least partially caused by $Climate$ and GDP. Thus, the variable $Institutions$ is a collider. If a researcher conditions on $Institutions$, this introduces collider bias. This figure is created with the tool *DAGitty* (Textor et al. 2016).

they condition on proxies for institutional quality and thereby only look at the relationship between countries with similar institutional strength. This could, for instance, happen if a researcher investigates this relationship in a sample that is only available for countries with a strong capacity for collecting statistics.[20]

Solutions for the replication crisis

The replication crisis can be tackled both by researchers individually and at the institutional level. For instance, research institutions and funders can increase incentives for researchers to run replication studies.[21] These replications may consist of a re-analysis of the same data or studies of a similar question in a different context. Beside replications, institutions could also create incentives for researchers to do meta-analyses, which have become a more common sight in interdisciplinary journals, but are still rare in economics. Particularly for interventions that we expect to have heterogeneous effects, running a meta-analysis can allow researchers to recover some "deep parameters", i.e. causal effects that hold across contexts. However, for meta-analyses, the old computer science principle "garbage in, garbage out" is valid, too; i.e., the quality of meta-analyses critically depends on the quality of the studies on which they are based. Well-identified studies are thus an ingredient of a credible meta-analysis. It is also useful for meta-analysis, if studies include all the specifications that have been run for a study, including the "insignificant" results. Authors should also correct their significance testing for multiple hypothesis testing if they explore the effects of a treatment on several outcomes and intend to report on these

[20]Collider bias can often be created by sample selection. For instance, when only looking at basketball players, height and performance could be less or even inversely correlated, as opposed to when looking at this relationship in the overall population. Being a basketball player is a collider here. It is caused by both the variables height and performance. See Schneider (2020) for additional examples of collider bias from economic history, which often deals with selected data. They revisit, for instance, the Acemoglu et al. (2001) paper that uses settler mortality as an instrument for the level of extractiveness of institutions. They proxy settler mortality with incomplete data, often collected long after colonisation. This data is, therefore, driven by institutions themselves, for data from the Caribbean and Latin America particularly, by the fact that Yellow Fever was brought there on slave ships from Africa (Albouy 2012). Thus, the proxy for settler mortality is both driven by the phenomenon of settler mortality itself, but also by the variable institutions introducing at least some level of collider bias.

[21]Note, however, that similar problems of cherry-picking can exist for replicators who try to reverse a result.

outcomes as individual findings. For instance, in Chapter 2 on the effects of palm oil investments, I correct the p-values of my more exploratory analysis of district outcomes for multiple hypothesis testing and indicate significance with an additional symbol beside the significance stars.

One of the most powerful tools to reduce both problems with researcher cherry-picking and publication bias are so called pre-analysis plans, documents detailing the most important elements of a research design, deposited with a trustworthy third actor before data is seen or collected. I discuss pre-registration and pre-analysis plans in the next Section.

Pre-analysis plans

Many journals prefer results that are precisely estimated and publish only a few papers that do not meet certain criteria for statistical significance (for instance a p-value of less than 0.05). This creates incentives for researchers to make conscious or unconscious choices in their analysis that lead them to overstate the statistical precision of their results, under-report less precise specifications or even report entirely spurious findings. A related issue is publication bias and, additionally, the preference of some journals for surprising and extreme findings. Since many incentives of researchers are based on publication, i.e., output rather than input, they may also take lower risks in their selection of research questions. *Pre-registration* of studies, in combination with pre-analysis plans, can help address some of these problems. Pre-registrations make sure that other researchers know that someone is undertaking a certain research project. This ensures that, even if a study is never written up and published, the project idea does not end up hidden entirely in a file drawer.

Pre-registrations are only the first step, since they only create transparency on a researcher's high-level research question, but do not necessarily include information on the exact research design. This is what pre-analysis plans do. In pre-analysis plans, researchers detail crucial elements of their research design before they see or collect the data (Miguel et al. 2014). This includes methods for data collection, data cleaning, and data analysis and lets researchers credibly distinguish in their papers those results based on a pre-specified protocol and those based on deviations from the plan. Some journals, such as the Journal of Development Economics have now even started to offer peer-review and an acceptance based on a pre-analysis plan only. Experimental economists have taken the lead with pre-analysis plans, but there is also an ongoing debate on their use for observational studies (Burlig 2018; Coffman and Niederle 2015; Olken 2015). Broadly, there are two camps: those who draw a clear line between pre-specified, confirmatory and exploratory analyses and those who suggest results should be assessed with robustness checks and sensitivity analyses, as advocated earlier by Leamer in his 1983 paper "Let's Take the Con Out of Econometrics", and further tested in replication studies.

Pre-analysis plans can even open up new, data-driven possibilities for causal inference. If

researchers have enough data, they can split their dataset in half to perform exploratory analyses on the first half, write up a pre-analysis plan before even seeing the second half of the data, and then run a confirmatory analysis on the set-aside data (Anderson and Magruder 2017; Fafchamps and Labonne 2017). Another approach in the context of the difference-in-differences design would be to explore pre-treatment data to find a counterfactual that matches the treatment group well without looking at the post-treatment data.[22] In a new pre-registered study on the German carbon price on fossil fuels (*Brennstoffemissionshandelsgesetzes*), my co-authors apply this strategy (Naumann et al. 2020). We use an extension to the synthetic control method (Ben-Michael et al. 2020) to construct a data-driven, synthetic counterfactual Germany and pre-register the study and the code behind this procedure before the policy becomes effective. This type of pre-registered synthetic control method could be a model for the evaluation of future policies. However, for interventions whose effect we only expect to build up over a period of years, this means an important investment by individual researchers long before they can measure the results and reap the benefits.

7.3 Conclusion

In this dissertation, I have built four plausibly unbiased counterfactuals. These counterfactuals incorporate the newest findings on the caveats of panel regression techniques. This research would not have been possible without new sources of data, for instance from satellites or from open web interfaces.

However, in this Synthesis and Outlook chapter, I have argued that one well-crafted, unbiased counterfactual is not enough to solve a research question. This is because findings from one context can rarely be easily generalized to other contexts or to the future. For some questions, such as those about mechanisms acting at an highly aggregate level, such as the EU-wide effect of carbon pricing on innovation, it may not even be possible to resolve the trade-off between bias and generalizability in a simple counterfactual. In these cases it may be most prudent to rely on more structure in a model. Directed acyclic graphs (DAGs) are one way to think more about the structure of a problem before estimating the "reduced-form". Bringing DAGs together with economic theory could be a fruitful way to improve our empirics on questions for which an identification strategy with plausibly random variation is not available.

One unbiased and generalizable counterfactual is, however, still not enough to solve a research question. The replication crisis has alerted us that studies should be replicable and that a meta-analysis based on solid studies may bring us closer to the truth than one paper published in a "Top Five Journal". For both systematic replications and meta-analysis,

[22]Synthetic control designs typically rely on a long pre-treatment panel to avoid overfitting (Abadie 2021). For difference-in-differences, Roth (2020) shows that if researchers select model specifications conditional on tests for pre-trends, bias can be worse than when using a model unconditionally on passing these tests.

well-structured and transparent code is essential. With ever larger datasets, applied economics has become a computer code-based team sport, but our practices often lag far behind those of computer scientists, who do pair-programming and code reviews, write tests and pseudo-code before they write code, and package coding environments so that others can run their code with a click of the mouse.[23] While some of the discussions on incentives in science have focused on malpractice, such as cherry-picking or even outright fraud, it may be even more important to make sure that researchers get institutional credit for providing a service to their discipline by making their work easily replicable and checking and replicating others' work.

However, a meta-analysis of replicated, fairly generalizable, unbiased counterfactuals would help us gauge the magnitude of a causal parameter conclusively – at least until the world changes and we have to start over.

[23]Economists working with econometric methods may be able to learn from their colleagues who use numerical models.

7.4 References

Abadie, Alberto (2021). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects". *Journal of Economic Literature*. DOI: 10.1257/jel.20191450.

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation". *American Economic Review* 91.5, pp. 1369–1401. DOI: 10.1257/aer.91.5.1369.

Albouy, David Y. (2012). "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment". *American Economic Review* 102.6, pp. 3059–3076. DOI: 10.1257/aer.102.6.3059.

Anderson, Michael L. and Jeremy Magruder (2017). *Split-Sample Strategies for Avoiding False Discoveries*. Working Paper 23544. National Bureau of Economic Research. DOI: 10.3386/w23544.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ashenfelter, Orley (1978). "Estimating the Effect of Training Programs on Earnings". *The Review of Economics and Statistics* 60.1, p. 47. DOI: 10.2307/1924332.

Athey, Susan and Guido Imbens (2018). *Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption*. Working Paper 24963. National Bureau of Economic Research. DOI: 10.3386/w24963.

Athey, Susan and Guido W. Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation". *Journal of Economic Perspectives* 31.2, pp. 3–32. DOI: 10.1257/jep.31.2.3.

Auffhammer, Maximilian (2018). "Quantifying Economic Damages from Climate Change". *Journal of Economic Perspectives* 32.4, pp. 33–52. DOI: 10.1257/jep.32.4.33.

Bellemare, Marc F., Takaaki Masaki, and Thomas B. Pepinsky (2017). "Lagged Explanatory Variables and the Estimation of Causal Effect". *The Journal of Politics* 79.3, pp. 949–963. DOI: 10.1086/690946.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2020). *The Augmented Synthetic Control Method*. URL: http://arxiv.org/abs/1811.04170 (visited on 01/23/2021).

Burlig, Fiona (2018). "Improving Transparency in Observational Social Science Research: A Pre-Analysis Plan Approach". *Economics Letters* 168, pp. 56–60. DOI: 10.1016/j.econlet.2018.03.036.

Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer, Adam Altmejd, Nick Buttrick, Taizan Chan, Yiling Chen, Eskil Forsell, Anup Gampa, Emma Heikensten, Lily Hummer, Taisuke Imai, Siri Isaksson, Dylan Manfredi, Julia Rose, Eric-Jan Wagenmakers, and Hang Wu (2018). "Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015". *Nature Human Behaviour* 2.9 (9), pp. 637–644. DOI: 10.1038/s41562-018-0399-z.

Cappelen, Alexander W., Odd-Helge Fjeldstad, Donald Mmari, Ingrid Hoem Sjursen, and Bertil Tungodden (2021). "Understanding the Resource Curse: A Large-Scale Experiment on Corruption in Tanzania". *Journal of Economic Behavior & Organization* 183, pp. 129–157. DOI: 10.1016/j.jebo.2020.12.027.

Coffman, Lucas C. and Muriel Niederle (2015). "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible". *Journal of Economic Perspectives* 29.3, pp. 81–98. DOI: 10.1257/jep.29.3.81.

Cust, James and Steven Poelhekke (2015). "The Local Economic Impacts of Natural Resource Extraction". *Annual Review of Resource Economics* 7.1, pp. 251–268. DOI: 10.1146/annurev-resource-100814-125106.

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. DOI: 10.1093/restud/rdz017.

Deschenes, Olivier and Kyle C. Meng (2018). "Chapter 7 - Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges". *Handbook of Environmental Economics*. Ed. by Partha Dasgupta, Subhrendu K. Pattanayak, and V. Kerry Smith. Vol. 4. Handbook of Environmental Economics. Elsevier, pp. 285–332. DOI: 10.1016/bs.hesenv.2018.08.001.

Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A. Nosek, and Magnus Johannesson (2015). "Using Prediction Markets to Estimate the Reproducibility of Scientific Research". *Proceedings of the National Academy of Sciences* 112.50, pp. 15343–15347. DOI: 10.1073/pnas.1516179112.

Duflo, Esther (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment". *American Economic Review* 91.4, pp. 795–813. DOI: 10.1257/aer.91.4.795.

Elwert, Felix and Christopher Winship (2014). "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable". *Annual Review of Sociology* 40.1, pp. 31–53. DOI: 10.1146/annurev-soc-071913-043455.

Fafchamps, Marcel and Julien Labonne (2017). "Using Split Samples to Improve Inference on Causal Effects". *Political Analysis* 25.4, pp. 465–482. DOI: 10.1017/pan.2017.22.

Gelman, Andrew and Eric Loken (2017). *The Statistical Crisis in Science*. American Scientist. URL: https://www.americanscientist.org/article/the-statistical-crisis-in-science (visited on 01/20/2021).

Goodman-Bacon, Andrew (2018). *Difference-in-Differences with Variation in Treatment Timing*. Working Paper 25018. National Bureau of Economic Research. DOI: 10.3386/w25018.

Gordon, Michael, Domenico Viganola, Michael Bishop, Yiling Chen, Anna Dreber, Brandon Goldfedder, Felix Holzmeister, Magnus Johannesson, Yang Liu, Charles Twardy, Juntao Wang, and Thomas Pfeiffer (2020). "Are Replication Rates the Same across Academic Fields? Community Forecasts from the DARPA SCORE Programme". *Royal Society Open Science* 7.7, p. 200566. DOI: 10.1098/rsos.200566.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods". *Econometrica* 37.3, pp. 424–438. DOI: 10.2307/1912791.

Hernán, Miguel A. and Stephen R. Cole (2009). "Invited Commentary: Causal Diagrams and Measurement Bias". *American Journal of Epidemiology* 170.8, pp. 959–962. DOI: 10.1093/aje/kwp293.

Hernán, Miguel A and James M Robins (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Hume, David (1748). *An Enquiry Concerning Human Understanding*. London.

Hünermund, Author Paul (2019). *Don't Put Too Much Meaning Into Control Variables*. Paul Hünermund. URL: https://p-hunermund.com/2019/04/28/dont-put-too-much-meaning-into-control-variables/ (visited on 01/22/2021).

Jedwab, Remi and Alexander Moradi (2015). "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa". *The Review of Economics and Statistics* 98.2, pp. 268–284. DOI: 10.1162/REST_a_00540.

Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kalkuhl, Matthias and Leonie Wenz (2020). "The Impact of Climate Conditions on Economic Production. Evidence from a Global Panel of Regions". *Journal of Environmental Economics and Management* 103, p. 102360. DOI: 10.1016/j.jeem.2020.102360.

Kolstad, Charles D. and Frances C. Moore (2020). "Estimating the Economic Impacts of Climate Change Using Weather Observations". *Review of Environmental Economics and Policy* 14.1, pp. 1–24. DOI: 10.1093/reep/rez024.

Leamer, Edward E. (1983). "Let's Take the Con Out of Econometrics". *The American Economic Review* 73.1, pp. 31–43.

McIntosh, Craig T and Wolfram Schlenker (2006). "Identifying Non-Linearities In Fixed Effects Models". Mimeo.

McKenzie, David (2020). *Revisiting the Difference-in-Differences Parallel Trends Assumption: Part I Pre-Trend Testing*. World Bank Blogs: Development Impact. URL: https://blogs.worldbank.org/impactevaluations/revisiting-difference-differences-parallel-trends-assumption-part-i-pre-trend (visited on 03/23/2021).

Menard, Alvaro de (2020). *What's Wrong with Social Science and How to Fix It: Reflections After Reading 2578 Papers*. Fantastic Anachronism. URL: https://fantasticanachronism.com/2020/09/11/whats-wrong-with-social-science-and-how-to-fix-it/index.html (visited on 01/20/2021).

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw (1994). "The Impact of Global Warming on Agriculture: A Ricardian Analysis". *The American Economic Review* 84.4, pp. 753–771.

Mérel, Pierre and Matthew Gammans (2021). "Climate Econometrics: Can the Panel Approach Account for Long-Run Adaptation?" *American Journal of Agricultural Economics* (forthcoming).

Meyfroidt, P, J Börner, R Garrett, T Gardner, J Godar, K Kis-Katos, B S Soares-Filho, and S Wunder (2020). "Focus on Leakage and Spillovers: Informing Land-Use Governance in a Tele-Coupled World". *Environmental Research Letters* 15.9, p. 090202. DOI: 10.1088/1748-9326/ab7397.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, and M. Van der Laan (2014). "Promoting Transparency in Social Science Research". *Science* 343.6166, pp. 30–31. DOI: 10.1126/science.1245317.

Naumann, Lennard, Nicolas Koch, and Sebastian Kraus (2020). *Evaluating the Effect of a CO_2 Price on Car Usage in Germany: An Augmented Synthetic Control Approach*. URL: https://osf.io/9ewj3 (visited on 01/19/2021).

Olken, Benjamin A. (2015). "Promises and Perils of Pre-Analysis Plans". *Journal of Economic Perspectives* 29.3, pp. 61–80. DOI: 10.1257/jep.29.3.61.

Pearl, Judea (2009). *Causality*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511803161.

Roessler, Philip, Yannick I. Pengl, Robert Marty, Kyle Sorlie Titlow, and Nicolas van de Walle (2020). *The Cash Crop Revolution, Colonialism and Legacies of Spatial Inequality: Evidence from Africa*. 2020-12. Centre for the Study of African Economies, University of Oxford.

Roth, Jonathan (2020). "Pre-Test with Caution: Event-Study Estimates After Testing for Parallel Trends". Mimeo.

Schneider, Eric B. (2020). "Collider Bias in Economic History Research". *Explorations in Economic History* 78, p. 101356. DOI: 10.1016/j.eeh.2020.101356.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant". *Psychological Science* 22.11, pp. 1359–1366. DOI: 10.1177/0956797611417632.

Textor, Johannes, Benito van der Zander, Mark S Gilthorpe, Maciej Liśkiewicz, and George TH Ellison (2016). "Robust Causal Inference Using Directed Acyclic Graphs: The R Package 'Dagitty'". *International Journal of Epidemiology* 45.6, pp. 1887–1894. DOI: 10.1093/ije/dyw341.

Van der Ploeg, Frederick (2011). "Natural Resources: Curse or Blessing?" *Journal of Economic Literature* 49.2, pp. 366–420. DOI: 10.1257/jel.49.2.366.

List of Tables

2.1 Number of treated and control districts for each cohort in the stacked dataset 49
2.2 Summary statistics manufacturing plants 53
2.3 Leads and lags of palm oil mill establishment on local manufacturing plant performance 56
2.4 Effects of palm oil mill establishment on local manufacturing plant performance 57
2.5 Effects on population, employment, and poverty at manufacturing plant and district levels 58
2.6 Effects on district budgets and infrastructure 60
2.7 Effects on manufacturing plant output portfolio 62
2.8 Checking SUTVA – Effects of palm oil mill establishment on neighboring districts 65
2.9 Corporate palm oil groups and their number of mills 79
2.10 Manufacturing census cleaning 80
2.11 TFP estimation: Production function coefficients by sub-sector 81
2.12 Effects of palm oil mill establishment on local manufacturing plant turnover . 82
2.13 Effects of palm oil mill establishment on downstream and upstream local manufacturing plant performance 82

3.1 Heterogeneous treatment effects of the pop-up bike lane roll-out 96
3.2 Summary statistics at the counter-day level 107
3.3 Summary statistics of most recent state of infrastructure at the city level . . . 107
3.4 Different treatment specifications 107
3.5 Additional robustness checks for our preferred estimate 108
3.6 Estimates of the average effect of pop-up bike lanes on cycling 109

4.1 Effect of community land titling on deforestation 122

5.1 Deforestation accumulated over 2002-2014, in kha 146
5.2 Estimation sample – descriptive statistics 157
5.3 Price elasticities of deforestation across Indonesian oil palm plantations . . . 159
5.4 Price elasticities of immediate and transitional deforestation in Indonesian industrial plantations 161
5.5 Palm fruit and crude palm oil price elasticities of deforestation across Indonesian oil palm plantations 163
5.6 Short-run and medium-run price elasticities of deforestation across Indonesian oil palm plantations 165

5.7 Counterfactual annual effects of different CPO price changes on deforestation in Indonesia . . . 168
5.8 IBS descriptive statistics . . . 177
5.9 Price elasticity and partial effects of control variables on deforestation across Indonesian oil palm sectors . . . 178
5.10 Price elasticities of deforestation across the oil palm sector, by island . . . 178
5.11 Price elasticity heterogeneity across ownership and local market development 179
5.12 p-values from equality tests of price elasticities . . . 179

6.1 The effect of thermal inversions on ambient air pollution . . . 197
6.2 The effect of ambient air pollution on COVID-19 outcomes . . . 198
6.3 The effect of thermal inversions on COVID-19 outcomes – full sample vs. lock-down sample . . . 200
6.4 Pollution effects on COVID-19 case numbers by days . . . 201
6.5 Pollution effects on COVID-19 deaths by days . . . 201
6.6 Overview of the COVID-19 data sources . . . 203
6.7 The effect of thermal inversions on movement . . . 205

List of Figures

1.1 Standard difference-in-differences design 15
1.2 Generalized difference-in-differences with a level-shift treatment 17
1.3 Decomposition of generalized (staggered) difference-in-differences with level-shift 18
1.4 Trend-break in staggered difference-in-differences 20
1.5 Constructing a stacked sample 22

2.1 Construction of the stacked dataset 41
2.2 Palm oil mills (location and establishment year) and number of manufacturing plants in bordering tracts with 1993 district boundaries 48
2.3 Distribution of palm oil mill establishment years in the palm oil mill panel . . 50
2.4 Dynamic effects of palm oil mill establishment on sales, labor productivity and total factor productivity 54
2.5 Coefficients from different specifications with sales (log) 63
2.6 Coefficients from different specifications with labor productivity (log) 83
2.7 Coefficients from different specifications with total factor productivity (log) . 84

3.1 Treated cities and their treatment intensities in terms of implemented kilometers of public bike lanes in service (cumulative) on a given day between March and July 2020 90
3.2 Treatment effect (difference between treated and control cities) in months before and after the beginning of the pop-up bike lane policy 91
3.3 Estimates of the average effect of pop-up bike lanes on cycling. Dose-response regressions (in km, km per capita, or km per km^2 in service on a given day) are multiplied by the average treatment dose 93
3.4 Intensity of pop-up bike lane treatment over time in treatment cities and control cities 110
3.5 Average bike count per week in treated cities 111
3.6 Average bike count per week in control cities 112
3.7 Change in daily cycling after the first day of treatment for each city 113
3.8 Treatment effect (difference between treated and control cities) in months before and after the beginning of the pop-up bike lane policy 113

4.1 Community title area locations and primary forest 118

5.1 Estimates of the Indonesian price elasticity of deforestation under different specifications 180

5.2 Estimates of price elasticity of land use change from forest to industrial plantations in Sumatra and Kalimantan under different specifications 181
5.3 Estimates of price elasticity of land use change from forest to industrial plantations in Sumatra under different specifications 182
5.4 Estimates of price elasticity of land use change from forest to industrial plantations in Kalimantan under different specifications 183
5.5 Estimates of price elasticity of land use change from forest to smallholder plantations in Sumatra under different specifications 184
5.6 Estimates of price elasticity of land use change from forest to industrial plantations in Kalimantan under different specifications 185

6.1 Spatial and temporal variation in the occurrence of inversion episodes 194
6.2 Alternative weather controls 199

7.1 Omitted variable bias from a backdoor in a directed acyclic graph (DAG) . . . 221
7.2 Irrigation is a mediator between climate and damages 222
7.3 Institutions are a collider variable for the relationship between climate and GDP 223

Tools and resources

This dissertation is written in LaTeX, both on the Overleaf platform and locally using Sublime Text 3.2.2 and TeX Live 2019/W32TeX.[24] Data cleaning and analysis are implemented with the following programming languages and environments: Stata MP (14.2), R (4.0.3), Python (2.7.15), and Google Earth Engine. Data sources are cited in the respective papers. Any data that can be shared openly, for instance because it is non-confidential and non-proprietary, has been included in the data and code packages cited in the individual papers.

Chapter 2

Data cleaning, analyses, tables, and figures in Stata MP (14.2) are implemented using native functions and the following packages:

`gtools`	(Bravo 2019)
`tsspell`	(Cox 2014)
`strgroup`	(Reif 2015)
`grcomb`	(Gamma 2010)
`fre`	(Jann 2015)
`distinct`	(Longton and Cox 2012)
`reghdfe`	(Correia 2015, 2016; Correia et al. 2020a)
`prodest`	(Rovigatti and Mollisi 2018, 2020)
`regsave`	(Reif 2020)
`estout`	(Jann 2007, 2019)

Stata charts are styled with the `plotplain` scheme (Bischof 2017). Code for *event study plots* is based on (Deshpande and Li 2019). Individual values and custom tables generated in Stata are transformed into LaTeXwith `stata-tex` (Novosad 2021).
Data cleaning and GIS functions in R are implemented using native functions and the following packages:

`dplyr`	(Wickham et al. 2021)
`plyr`	(Wickham 2011, 2020c)
`tidyr`	(Wickham 2021)
`tibble`	(Müller and Wickham 2020)
`readxl`	(Wickham and Bryan 2019)
`xlsx`	(Dragulescu and Arendt 2020)
`data.table`	(Dowle and Srinivasan 2021)
`haven`	(Wickham and Miller 2020)
`foreign`	(R Core Team 2020)
`readstata13`	(Garbuszus and Jeworutzki 2018)
`rio`	(Chan and Leeper 2021)
`stringr`	(Wickham 2019b)
`stringdist`	(van der Loo 2014; van der Loo 2020)
`raster`	(Hijmans 2020)
`sf`	(Pebesma 2018, 2021)
`sp`	(Bivand et al. 2013a; Pebesma and Bivand 2005; Pebesma and Bivand 2021)
`spdep`	(Bivand 2021; Bivand et al. 2013b; Bivand and Wong 2018)
`rgdal`	(Bivand et al. 2021)
`exactextractr`	(Daniel Baston 2020)
`geosphere`	(Hijmans 2019)
`mapdata`	(Richard A. Becker and Ray Brownrigg. 2018)
`tmap`	(Tennekes 2021)
`leaflet`	(Cheng et al. 2021a)
`htmltools`	(Cheng et al. 2021b)
`eeptools`	(Knowles 2020)
`GISTools`	(Brunsdon and Chen 2014)
`RColorBrewer`	(Neuwirth 2014)
`ggplot2`	(Wickham et al. 2020a)

[24] A tutorial on using Sublime Text as a text editor for LaTeX can be found at https://acarril.github.io/posts/use-st3

Specification charts are built with code by Ortiz-Bobea (2020).

Chapter 3

Data cleaning, analyses, tables, and figures in Stata MP (14.2) are implemented using native functions and the following software packages:

`gtools`	(Bravo 2019)	`ppmlhdfe`	(Correia et al. 2019, 2020b)
`strgroup`	(Reif 2015)	`regsave`	(Reif 2020)
`reghdfe`	(Correia 2015, 2016; Correia et al. 2020a)	`estout`	(Jann 2007, 2019)

Stata charts are styled with the `plotplain` scheme (Bischof 2017).
Data cleaning, GIS functions, and analyses in R (4.0.3) are implemented using native functions and the following packages:

`dplyr`	(Wickham et al. 2021)	`ggthemes`	(Arnold 2019)
`tidyr`	(Wickham 2021)	`viridis`	(Garnier 2018a)
`tidyverse`	(Wickham 2019c)	`viridisLite`	(Garnier 2018b)
`tibble`	(Müller and Wickham 2020)	`extrafont`	(Chang 2014)
`readr`	(Wickham et al. 2018)	`scales`	(Wickham and Seidel 2020)
`xlsx`	(Dragulescu and Arendt 2020)	`ecmwfr`	(Hufkens 2020; Hufkens et al. 2019)
`haven`	(Wickham and Miller 2020)	`raster`	(Hijmans 2020)
`jsonlite`	(Ooms 2014, 2020a)	`sf`	(Pebesma 2018, 2021)
`rjson`	(Couture-Beil 2018)	`sp`	(Bivand et al. 2013a; Pebesma and Bivand 2005; Pebesma and Bivand 2021)
`rlist`	(Ren 2016)	`geojsonsf`	(Cooley 2020)
`lubridate`	(Grolemund and Wickham 2011; Spinu et al. 2021)	`exactextractr`	(Daniel Baston 2020)
`forcats`	(Wickham 2020a)	`panelView`	(Liu and Xu 2020)
`janitor`	(Firke 2020)	`fect`	(Liu et al. 2020)
`purrr`	(Henry and Wickham 2020)	`augsynth`	(Ben-Michael 2021)
`stringr`	(Wickham 2019b)	`gsynth`	(Xu and Liu 2020)
`ggplot2`	(Wickham 2016; Wickham et al. 2020a)		

Chapter 4

Data cleaning, analyses, tables, and figures in Stata MP (14.2) are implemented using native functions and the following software packages:

`gtools`	(Bravo 2019)	`ppmlhdfe`	(Correia et al. 2019, 2020b)
`reghdfe`	(Correia 2015, 2016; Correia et al. 2020a)	`regsave`	(Reif 2020)
		`estout`	(Jann 2007, 2019)

Stata charts are styled with the `plotplain` scheme (Bischof 2017).
Data cleaning, GIS functions, and analyses in R (4.0.3) are implemented using native functions and the following packages:

abind	(Plate and Heiberger 2016)	fasterize	(Ross 2020)
dplyr	(Wickham et al. 2021)	rgeos	(Ross 2020)
haven	(Wickham and Miller 2020)	sf	(Pebesma 2018, 2021)
readr	(Wickham et al. 2018)	sp	(Bivand et al. 2013a; Pebesma and Bivand 2005; Pebesma and Bivand 2021)
exactextractr	(Daniel Baston 2020)		

Deforestation data is matched to study areas directly on Google Earth Engine (Gorelick et al. 2017).
Maps are prepared in R (4.0.3) and styled in QGIS (3.2.3).

Chapter 5

Data cleaning, GIS functions, and analyses in R (4.0.3) are implemented using native functions and the following packages:

emmeans	(Lenth 2021)	osrmr	(Staempfli and Strauss 2019)
ggeffects	(Lüdecke 2021a)	osrm	(Giraud 2021)
htmltools	(Cheng et al. 2021b)	gfcanalysis	(Zvoleff 2020)
leaflet	(Cheng et al. 2021a)	spdep	(Bivand 2021)
broom	(Robinson et al. 2021)	sf	(Pebesma 2021)
lmtest	(Hothorn et al. 2020)	spData	(Bivand et al. 2020)
zoo	(Zeileis et al. 2021)	rgdal	(Bivand et al. 2021)
sandwich	(Zeileis and Lumley 2020)	raster	(Hijmans 2020)
fixest	(Berge 2021)	sp	(Pebesma and Bivand 2021)
car	(Fox et al. 2020a)	haven	(Wickham and Miller 2020)
carData	(Fox et al. 2020b)	httr	(Wickham 2020b)
multcomp	(Hothorn et al. 2021)	writexl	(Ooms 2020b)
TH.data	(Hothorn 2019)	readxl	(Wickham and Bryan 2019)
MASS	(Ripley 2020)	foreign	(R Core Team 2020)
mvtnorm	(Genz et al. 2020)	readstata13	(Garbuszus and Jeworutzki 2018)
DataCombine	(Gandrud 2016)	stringr	(Wickham 2019b)
kableExtra	(Zhu 2021)	sjmisc	(Lüdecke 2021b)
knitr	(Xie 2021)	Hmisc	(Harrell 2021)
snow	(Tierney et al. 2018)	ggplot2	(Wickham et al. 2020a)
doParallel	(Corporation and Weston 2020)	Formula	(Zeileis and Croissant 2020)
iterators	(Analytics and Weston 2020)	survival	(Therneau 2020)
foreach	(Revolution Analytics and Weston n.d.)	lattice	(Sarkar 2020)
exactextractr	(Daniel Baston 2020)	dplyr	(Wickham et al. 2021)
lubridate	(Grolemund and Wickham 2011; Spinu et al. 2021)	tidyr	(Wickham 2021)
nngeo	(Dorman 2021)	plyr	(Wickham 2020c)
		data.table	(Dowle and Srinivasan 2021)

Chapter 6

Data cleaning, GIS functions, and analyses in R (4.0.3) are implemented using native functions and the following packages:

`rsample`	(Kuhn et al. 2020)
`MMWRweek`	(Niemi 2020)
`rlist`	(Ren 2016)
`ecmwfr`	(Hufkens 2020)
`qdapRegex`	(Rinker 2017a; Rinker 2017b)
`httr`	(Wickham 2020b)
`rNOMADS`	(Bowman 2020; Bowman and Lees 2015)
`rvest`	(Wickham 2019a)
`xml2`	(Wickham et al. 2020b)
`pglm`	(Croissant 2020)
`plm`	(Croissant and Millo 2008, 2018; Croissant et al. 2020; Millo 2017)
`maxLik`	(Henningsen and Toomet 2011; Toomet et al. 2020)
`miscTools`	(Henningsen and Toomet 2019)
`RcppRoll`	(Ushey 2018)
`plyr`	(Wickham 2011, 2020c)
`reshape2`	(Wickham 2007, 2020d)
`stringdist`	(van der Loo 2014; van der Loo 2020)
`data.table`	(Dowle and Srinivasan 2021)
`rjson`	(Couture-Beil 2018)
`stringr`	(Wickham 2019b)
`haven`	(Wickham and Miller 2020)
`mice`	(van Buuren and Groothuis-Oudshoorn 2011, 2021)
`tiff`	(Urbanek and Johnson 2021)
`readr`	(Wickham et al. 2018)
`worldmet`	(Carslaw 2020)
`tidyr`	(Wickham 2021)
`zoo`	(Zeileis and Grothendieck 2005; Zeileis et al. 2021)
`car`	(Fox et al. 2020a)
`carData`	(Fox et al. 2020b)
`solartime`	(Wutzler 2018)
`lutz`	(Teucher 2019)
`future`	(Bengtsson 2020)
`reshape`	(Wickham 2018)
`ncdf4`	(Pierce 2019)
`sf`	(Pebesma 2018, 2021)
`raster`	(Hijmans 2020)
`sp`	(Bivand et al. 2013a; Pebesma and Bivand 2021)
`exactextractr`	(Daniel Baston 2020)
`fixest`	(Berge 2021)
`xtable`	(Dahl et al. 2019)
`purrr`	(Henry and Wickham 2020)
`lfe`	(Gaure 2013a,b, 2014, 2019, 2020)
`Matrix`	(Bates and Maechler 2019)
`broom`	(Robinson et al. 2021)
`lubridate`	(Grolemund and Wickham 2011; Spinu et al. 2021)
`dplyr`	(Wickham et al. 2021)
`estimatr`	(Blair et al. 2020)

References

Analytics, Revolution and Steve Weston (2020). *iterators: Provides Iterator Construct*. R package version 1.0.13.

Arnold, Jeffrey B. (2019). *ggthemes: Extra Themes, Scales and Geoms for ggplot2*. R package version 4.2.0.

Bates, Douglas and Martin Maechler (2019). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-18.

Ben-Michael, Eli (2021). *augsynth: The Augmented Synthetic Control Method*. R package version 0.2.0.

Bengtsson, Henrik (2020). *future: Unified Parallel and Distributed Processing in R for Everyone*. R package version 1.18.0.

Berge, Laurent (2021). *fixest: Fast Fixed-Effects Estimations*. R package version 0.8.4.

Bischof, Daniel (2017). "New Graphic Schemes for Stata: Plotplain and Plottig". *The Stata Journal* 17.3, pp. 748–759. DOI: 10.1177/1536867X1701700313.

Bivand, Roger (2021). *spdep: Spatial Dependence: Weighting Schemes, Statistics*. R package version 1.1-7.

Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio (2013a). *Applied spatial data analysis with R, Second edition*. Springer, NY.

– (2013b). *Applied spatial data analysis with R, Second edition*. Springer, NY.

Bivand, Roger, Tim Keitt, and Barry Rowlingson (2021). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 1.5-23.

Bivand, Roger, Jakub Nowosad, and Robin Lovelace (2020). *spData: Datasets for Spatial Analysis*. R package version 0.3.8.

Bivand, Roger and David W. S. Wong (2018). "Comparing implementations of global and local indicators of spatial association". *TEST* 27.3, pp. 716–748.

Blair, Graeme, Jasper Cooper, Alexander Coppock, Macartan Humphreys, and Luke Sonnet (2020). *estimatr: Fast Estimators for Design-Based Inference*. R package version 0.22.0.

Bowman, Daniel C. (2020). *rNOMADS: An R Interface to the NOAA Operational Model Archive and Distribution System*. R package version 2.5.0.

Bowman, Daniel C. and Jonathan M. Lees (2015). "Near real time weather and ocean model data access with rNOMADS". *Computers & Geosciences* 78, pp. 88–95.

Bravo, Mauricio Caceres (2019). *GTOOLS: Stata Module to Provide a Fast Implementation of Common Group Commands*. Statistical Software Components.

Brunsdon, Chris and Hongyan Chen (2014). *GISTools: Some further GIS capabilities for R*. R package version 0.7-4.

Carslaw, David (2020). *worldmet: Import Surface Meteorological Data from NOAA Integrated Surface Database (ISD)*. R package version 0.9.0.

Chan, Chung-hong and Thomas J. Leeper (2021). *rio: A Swiss-Army Knife for Data I/O*. R package version 0.5.26.

Chang, Winston (2014). *extrafont: Tools for using fonts*. R package version 0.17.

Cheng, Joe, Bhaskar Karambelkar, and Yihui Xie (2021a). *leaflet: Create Interactive Web Maps with the JavaScript Leaflet Library*. R package version 2.0.4.1.

Cheng, Joe, Carson Sievert, Winston Chang, Yihui Xie, and Jeff Allen (2021b). *htmltools: Tools for HTML*. R package version 0.5.1.1.

Cooley, David (2020). *geojsonsf: GeoJSON to Simple Feature Converter*. R package version 2.0.0.

Corporation, Microsoft and Steve Weston (2020). *doParallel: Foreach Parallel Adaptor for the parallel Package*. R package version 1.0.16.

Correia, Sergio (2015). "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix", p. 7.

Correia, Sergio (2016). "A Feasible Estimator for Linear Models with Multi-Way Fixed Effects", p. 19.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin (2019). *Verifying the Existence of Maximum Likelihood Estimates for Generalized Linear Models*. Preprint.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin (2020a). "Fast Poisson Estimation with High-Dimensional Fixed Effects". *The Stata Journal* 20.1, pp. 95–115. DOI: 10.1177/1536867X20909691.

– (2020b). "Fast Poisson Estimation with High-Dimensional Fixed Effects". *The Stata Journal* 20.1, pp. 95–115. DOI: 10.1177/1536867X20909691.

Couture-Beil, Alex (2018). *rjson: JSON for R*. R package version 0.2.20.

Cox, Nicholas (2014). *TSSPELL: Stata Module for Identification of Spells or Runs in Time Series*. Statistical Software Components.

Croissant, Yves (2020). *pglm: Panel Generalized Linear Models*. R package version 0.2-2.

Croissant, Yves and Giovanni Millo (2008). "Panel Data Econometrics in R: The plm Package". *Journal of Statistical Software* 27.2, pp. 1–43.

– (2018). *Panel Data Econometrics with R: the plm package*. Wiley.

Croissant, Yves, Giovanni Millo, and Kevin Tappe (2020). *plm: Linear Models for Panel Data*. R package version 2.2-3.

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Daniel Baston (2020). *exactextractr: Fast Extraction from Raster Datasets using Polygons*. R package version 0.5.1.

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. DOI: 10.1257/pol.20180076.

Dorman, Michael (2021). *nngeo: k-Nearest Neighbor Join for Spatial Data*. R package version 0.4.2.

Dowle, Matt and Arun Srinivasan (2021). *data.table: Extension of 'data.frame'*. R package version 1.14.0.

Dragulescu, Adrian and Cole Arendt (2020). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. R package version 0.6.3.

Firke, Sam (2020). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. R package version 2.0.1.

Fox, John, Sanford Weisberg, and Brad Price (2020a). *car: Companion to Applied Regression*. R package version 3.0-10.

– (2020b). *carData: Companion to Applied Regression Data Sets*. R package version 3.0-4.

Gamma, Alex (2010). *GRCOMB: Stata Module to Create and Combine Several Single Graphs into One*. Statistical Software Components.

Gandrud, Christopher (2016). *DataCombine: Tools for Easily Combining and Cleaning Data Sets*. R package version 0.2.21.

Garbuszus, Jan Marvin and Sebastian Jeworutzki (2018). *readstata13: Import Stata Data Files*. R package version 0.9.2.

Garnier, Simon (2018a). *viridis: Default Color Maps from matplotlib*. R package version 0.5.1.

– (2018b). *viridisLite: Default Color Maps from matplotlib (Lite Version)*. R package version 0.3.0.

Gaure, Simen (2013a). "lfe: Linear group fixed effects". *The R Journal* 5.2. User documentation of the 'lfe' package, pp. 104–117.

– (2013b). "OLS with multiple high dimensional category variables". *Computational Statistics & Data Analysis* 66. Description of the projection methods used in 'lfe', pp. 8–18.

– (2014). "Correlation bias correction in two-way fixed effects linear regression". *Stat* 3.1. Description of the limited mobility bias correction method used in 'lfe', pp. 379–390.

– (2019). *lfe: Linear Group Fixed Effects*.

– (2020). *lfe: Linear Group Fixed Effects*. R package version 2.8-5.1.

Genz, Alan, Frank Bretz, Tetsuhisa Miwa, Xuefei Mi, and Torsten Hothorn (2020). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.1-1.

Giraud, Timothée (2021). *osrm: Interface Between R and the OpenStreetMap-Based Routing Service OSRM*. R package version 3.4.1.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". *Remote Sensing of Environment*. DOI: 10.1016/j.rse.2017.06.031.

Grolemund, Garrett and Hadley Wickham (2011). "Dates and Times Made Easy with lubridate". *Journal of Statistical Software* 40.3, pp. 1–25.

Harrell Jr., Frank E (2021). *Hmisc: Harrell Miscellaneous*. R package version 4.5-0.

Henningsen, Arne and Ott Toomet (2011). "maxLik: A package for maximum likelihood estimation in R". *Computational Statistics* 26.3, pp. 443–458.

– (2019). *miscTools: Miscellaneous Tools and Utilities*. R package version 0.6-26.

Henry, Lionel and Hadley Wickham (2020). *purrr: Functional Programming Tools*. R package version 0.3.4.

Hijmans, Robert J. (2019). *geosphere: Spherical Trigonometry*. R package version 1.5-10.

– (2020). *raster: Geographic Data Analysis and Modeling*. R package version 3.4-5.

Hothorn, Torsten (2019). *TH.data: TH's Data Archive*. R package version 1.0-10.

Hothorn, Torsten, Frank Bretz, and Peter Westfall (2021). *multcomp: Simultaneous Inference in General Parametric Models*. R package version 1.4-16.

Hothorn, Torsten, Achim Zeileis, Richard W. Farebrother, and Clint Cummins (2020). *lmtest: Testing Linear Regression Models*. R package version 0.9-38.

Hufkens, Koen (2020). *ecmwfr: Interface to ECMWF and CDS Data Web Services*. R package version 1.2.3.

Hufkens, Koen, Reto Stauffer, and Elio Campitelli (2019). *The ecmwfr package: an interface to ECMWF API endpoints*. DOI: 10.5281/zenodo.2647541.

Jann, Ben (2007). "Making Regression Tables Simplified". *The Stata Journal* 7.2, pp. 227–244. DOI: 10.1177/1536867X0700700207.

– (2015). *FRE: Stata Module to Display One-Way Frequency Table*. Statistical Software Components.

– (2019). *ESTOUT: Stata Module to Make Regression Tables*. Statistical Software Components.

Knowles, Jared E. (2020). *eeptools: Convenience Functions for Education Data*. R package version 1.2.4.

Kuhn, Max, Fanny Chow, and Hadley Wickham (2020). *rsample: General Resampling Infrastructure*. R package version 0.0.7.

Lenth, Russell V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.5.5-1.

Liu, Licheng, Ye Wang, Yiqing Xu, and Ziyi Liu (2020). *fect: Fixed Effects Counterfactuals*. R package version 0.3.1.

Liu, Licheng and Yiqing Xu (2020). *panelView: Visualizing Panel Data*. R package version 1.1.3.

Longton, Gary and Nicholas Cox (2012). *DISTINCT: Stata Module to Display Distinct Values of Variables*. Statistical Software Components.

Lüdecke, Daniel (2021a). *ggeffects: Create Tidy Data Frames of Marginal Effects for ggplot from Model Outputs*. R package version 1.0.2.

– (2021b). *sjmisc: Data and Variable Transformation Functions*. R package version 2.8.6.

Millo, Giovanni (2017). "Robust Standard Error Estimators for Panel Models: A Unifying Approach". *Journal of Statistical Software* 82.3, pp. 1–27.

Müller, Kirill and Hadley Wickham (2020). *tibble: Simple Data Frames*. R package version 3.0.1.

Neuwirth, Erich (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2.

Niemi, Jarad (2020). *MMWRweek: Convert Dates to MMWR Day, Week, and Year*. R package version 0.1.3.

Novosad, Paul (2021). *Stata-Tex*.

Ooms, Jeroen (2014). "The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects". *arXiv:1403.2805 [stat.CO]*.

– (2020a). *jsonlite: A Simple and Robust JSON Parser and Generator for R*. R package version 1.7.2.

– (2020b). *writexl: Export Data Frames to Excel xlsx Format*. R package version 1.3.1.

Ortiz-Bobea, Ariel (2020). *ArielOrtizBobea/Spec_chart*. https://github.com/ArielOrtizBobea/spec_chart.

Pebesma, Edzer (2018). "Simple Features for R: Standardized Support for Spatial Vector Data". *The R Journal* 10.1, pp. 439–446. DOI: 10.32614/RJ-2018-009.

– (2021). *sf: Simple Features for R*. R package version 0.9-8.

Pebesma, Edzer J. and Roger S. Bivand (2005). "Classes and methods for spatial data in R". *R News* 5.2, pp. 9–13.

Pebesma, Edzer and Roger Bivand (2021). *sp: Classes and Methods for Spatial Data*. R package version 1.4-5.

Pierce, David (2019). *ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files*. R package version 1.17.

Plate, Tony and Richard Heiberger (2016). *abind: Combine Multidimensional Arrays*. R package version 1.4-5.

R Core Team (2020). *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, . . .* R package version 0.8-80.

Reif, Julian (2015). *STRGROUP: Stata Module to Match Strings Based on Their Levenshtein Edit Distance*. Statistical Software Components.

– (2020). *REGSAVE: Stata Module to Save Regression Results to a Stata-Formatted Dataset*. Statistical Software Components.

Ren, Kun (2016). *rlist: A Toolbox for Non-Tabular Data Manipulation*. R package version 0.4.6.1.

Revolution Analytics and Steve Weston (n.d.). *foreach: Provides Foreach Looping Construct*.

Richard A. Becker, Original S code by and Allan R. Wilks. R version by Ray Brownrigg. (2018). *mapdata: Extra Map Databases*. R package version 2.3.0.

Rinker, Tyler (2017a). *qdapRegex: Regular Expression Removal, Extraction, and Replacement Tools*. R package version 0.7.2.

Rinker, Tyler W. (2017b). *qdapRegex: Regular Expression Removal, Extraction, and Replacement Tools*. 0.7.2. University at Buffalo/SUNY. Buffalo, New York.

Ripley, Brian (2020). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-53.

Robinson, David, Alex Hayes, and Simon Couch (2021). *broom: Convert Statistical Objects into Tidy Tibbles*. R package version 0.7.6.

Ross, Noam (2020). *fasterize: Fast Polygon to Raster Conversion*. R package version 1.0.2.

Rovigatti, Gabriele and Vincenzo Mollisi (2018). "Theory and Practice of Total-Factor Productivity Estimation: The Control Function Approach Using Stata". *The Stata Journal: Promoting communications on statistics and Stata* 18.3, pp. 618–662. DOI: 10.1177/1536867X1801800307.

– (2020). *PRODEST: Stata Module for Production Function Estimation Based on the Control Function Approach*. Statistical Software Components.

Sarkar, Deepayan (2020). *lattice: Trellis Graphics for R*. R package version 0.20-41.

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham (2021). *lubridate: Make Dealing with Dates a Little Easier*. R package version 1.7.10.

Staempfli, Adrian and Christoph Strauss (2019). *osrmr: Wrapper for the OSRM API*. R package version 0.1.35.

Tennekes, Martijn (2021). *tmap: Thematic Maps*. R package version 3.3-1.

Teucher, Andy (2019). *lutz: Look Up Time Zones of Point Coordinates*. R package version 0.3.1.

Therneau, Terry M (2020). *survival: Survival Analysis*. R package version 3.2-7.

Tierney, Luke, A. J. Rossini, Na Li, and H. Sevcikova (2018). *snow: Simple Network of Workstations*. R package version 0.4-3.

Toomet, Ott, Arne Henningsen, with contributions from Spencer Graves, and Yves Croissant (2020). *maxLik: Maximum Likelihood Estimation and Related Tools*. R package version 1.3-8.

Urbanek, Simon and Kent Johnson (2021). *tiff: Read and Write TIFF Images*. R package version 0.1-8.

Ushey, Kevin (2018). *RcppRoll: Efficient Rolling / Windowed Operations*. R package version 0.3.0.

van Buuren, Stef and Karin Groothuis-Oudshoorn (2011). "mice: Multivariate Imputation by Chained Equations in R". *Journal of Statistical Software* 45.3, pp. 1–67.

– (2021). *mice: Multivariate Imputation by Chained Equations*. R package version 3.13.0.

van der Loo, M.P.J. (2014). "The stringdist package for approximate string matching". *The R Journal* 6 (1), pp. 111–122.

van der Loo, Mark (2020). *stringdist: Approximate String Matching, Fuzzy Text Search, and String Distance Functions*. R package version 0.9.6.3.

Wickham, Hadley (2007). "Reshaping Data with the reshape Package". *Journal of Statistical Software* 21.12, pp. 1–20.

– (2011). "The Split-Apply-Combine Strategy for Data Analysis". *Journal of Statistical Software* 40.1, pp. 1–29.

– (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

– (2018). *reshape: Flexibly Reshape Data*. R package version 0.8.8.

– (2019a). *rvest: Easily Harvest (Scrape) Web Pages*. R package version 0.3.5.

– (2019b). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0.

– (2019c). *tidyverse: Easily Install and Load the Tidyverse*. R package version 1.3.0.

– (2020a). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 0.5.0.

– (2020b). *httr: Tools for Working with URLs and HTTP*. R package version 1.4.2.

– (2020c). *plyr: Tools for Splitting, Applying and Combining Data*. R package version 1.8.6.

– (2020d). *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. R package version 1.4.4.

– (2021). *tidyr: Tidy Messy Data*. R package version 1.1.3.

Wickham, Hadley and Jennifer Bryan (2019). *readxl: Read Excel Files*. R package version 1.3.1.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington (2020a). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.3.3.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller (2021). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.5.

Wickham, Hadley, Jim Hester, and Romain Francois (2018). *readr: Read Rectangular Text Data*. R package version 1.3.1.

Wickham, Hadley, Jim Hester, and Jeroen Ooms (2020b). *xml2: Parse XML*. R package version 1.3.2.

Wickham, Hadley and Evan Miller (2020). *haven: Import and Export SPSS, Stata and SAS Files*. R package version 2.3.1.

Wickham, Hadley and Dana Seidel (2020). *scales: Scale Functions for Visualization*. R package version 1.1.1.

Wutzler, Thomas (2018). *solartime: Utilities Dealing with Solar Time Such as Sun Position and Time of Sunrise*. R package version 0.0.1.

Xie, Yihui (2021). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.32.

Xu, Yiqing and Licheng Liu (2020). *gsynth: Generalized Synthetic Control Method*. R package version 1.1.7.

Zeileis, Achim and Yves Croissant (2020). *Formula: Extended Model Formulas*. R package version 1.2-4.

Zeileis, Achim and Gabor Grothendieck (2005). "zoo: S3 Infrastructure for Regular and Irregular Time Series". *Journal of Statistical Software* 14.6, pp. 1–27. DOI: 10.18637/jss.v014.i06.

Zeileis, Achim, Gabor Grothendieck, and Jeffrey A. Ryan (2021). *zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)*. R package version 1.8-9.

Zeileis, Achim and Thomas Lumley (2020). *sandwich: Robust Covariance Matrix Estimators*. R package version 3.0-0.

Zhu, Hao (2021). *kableExtra: Construct Complex Table with kable and Pipe Syntax*. R package version 1.3.4.

Zvoleff, Alex (2020). *gfcanalysis: Tools for Working with Hansen et al. Global Forest Change Dataset*. R package version 1.6.0.

Statement of contributions

The individual chapters in this dissertation are based on collaborations with co-authors. The Introduction (Chapter 1) and Conclusion (Chapter 7) are solo-authored.

Chapter 2 *Sebastian Kraus* is the lead author of this Chapter.[25] Sebastian Kraus designed the research in collaboration with Nicolas Koch and Robert Heilmayr. Sebastian Kraus cleaned and constructed the estimation sample with initial research assistance, conducted the analysis, designed figures and tables, and wrote the original draft. Sebastian Kraus, Nicolas Koch, and Robert Heilmayr reviewed and edited the paper.

Chapter 3 *Sebastian Kraus* is the lead author of this Chapter. Both authors designed the analysis, interpreted results, designed figures and wrote the paper. Sebastian Kraus acquired and cleaned the data with research assistance and conducted the analyses.[26]

Chapter 4 *Sebastian Kraus* is the lead author of this Chapter.[27] Sebastian Kraus and Jacqueline Liu conducted the analysis. Sebastian Kraus drafted the paper. All authors designed research, interpreted results and edited the paper.

Chapter 5 *Valentin Guye* is the lead author of this Chapter.[28] Valentin Guye designed the research in collaboration with Sebastian Kraus and with support from their respective supervisors. Sebastian Kraus cleaned the baseline data (manufacturing census). Valentin Guye cleaned and constructed the estimation sample, conducted the analysis, designed figures and tables, and wrote the original draft. Valentin Guye and Sebastian Kraus reviewed and edited the paper.

Chapter 6 *All authors* designed the research, interpreted results, designed figures and wrote this Chapter.[29] The authors contributed equally and are listed alphabetically.

[25]Sebastian Kraus, Robert Heilmayr, and Nicolas Koch (2020). "Spillovers to Manufacturing Plants from Multi-Million Dollar Plantations: Evidence from the Indonesian Palm Oil Boom". Revise and Resubmit *Journal of the Association of Environmental and Resource Economists* on 23 August 2021.

[26]Sebastian Kraus and Nicolas Koch (2021). "Provisional COVID-19 Infrastructure Induces Large, Rapid Increases in Cycling". *Proceedings of the National Academy of Sciences* 118.15. CC BY-NC-ND 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0/. DOI: 10.1073/pnas.2024399118

[27]Sebastian Kraus, Jacqueline Liu, Nicolas Koch, and Sabine Fuss (2021). "No Aggregate Deforestation Reductions from Rollout of Community Land Titles in Indonesia Yet". *Proceedings of the National Academy of Sciences* 118.43. CC BY 4.0 https://creativecommons.org/licenses/by/4.0/. DOI: 10.1073/pnas.2100741118

[28]Valentin Guye and Sebastian Kraus (2021). "Price incentives and unmonitored deforestation: Evidence from Indonesian palm oil mills". In preparation for resubmission at the *Journal of Environmental Economics and Management*.

[29]Hannah Klauber, Nicolas Koch, and Sebastian Kraus (2021). "Cross-country evidence on the effect of air pollution on COVID-19 from thermal inversions as a natural experiment". Under Review at *Environmental*

Hannah Klauber and Sebastian Kraus collected and cleaned the data. Hannah Klauber ran most of the analyzes with contributions by Sebastian Kraus.

Research Letters

List of publications

Chapter 2 *Preprint* Sebastian Kraus, Robert Heilmayr, and Nicolas Koch (2020). "Spillovers to Manufacturing Plants from Multi-Million Dollar Plantations: Evidence from the Indonesian Palm Oil Boom". Revise and Resubmit *Journal of the Association of Environmental and Resource Economists* on 23 August 2021.

Chapter 3 *Preprint* Sebastian Kraus and Nicolas Koch (2021). "Provisional COVID-19 Infrastructure Induces Large, Rapid Increases in Cycling". *Proceedings of the National Academy of Sciences* 118.15. CC BY-NC-ND 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0/. DOI: 10.1073/pnas.2024399118

Chapter 4 *Preprint* Sebastian Kraus, Jacqueline Liu, Nicolas Koch, and Sabine Fuss (2021). "No Aggregate Deforestation Reductions from Rollout of Community Land Titles in Indonesia Yet". *Proceedings of the National Academy of Sciences* 118.43. CC BY 4.0 https://creativecommons.org/licenses/by/4.0/. DOI: 10.1073/pnas.2100741118

Chapter 5 *Preprint* Valentin Guye and Sebastian Kraus (2021). "Price incentives and unmonitored deforestation: Evidence from Indonesian palm oil mills". In preparation for resubmission at the *Journal of Environmental Economics and Management*.

Chapter 6 *Preprint* Hannah Klauber, Nicolas Koch, and Sebastian Kraus (2021). "Cross-country evidence on the effect of air pollution on COVID-19 from thermal inversions as a natural experiment". Under Review at *Environmental Research Letters*

Acknowledgements

Working on this dissertation has been a privilege. I have learnt a lot about econometrics, data, computers, the scientific process, the nature of knowledge, and about myself. I am grateful to all the people who helped me on the way and to all those who were patient with me during this time of intense, focused, and often solitary work.

I would like to thank Ottmar Edenhofer, Sabine Fuss, and Nicolas Koch for supervising this dissertation and Robert Heilmayr for acting as a second referee.

I am deeply grateful for the opportunity to work at the Mercator Research Institute on Global Commons and Climate Change during my PhD. Ottmar Edenhofer and his team have created a unique space for learning, discussions, questioning, and pursuing new projects – even some risky and time-intensive ones. I cannot imagine a more stimulating environment to study climate and sustainability and I am indebted to all the people who make this place. Thanks to Brigitte Knopf, Elisabeth Nierhoff, Ulrich von Lampe, Susann Reinsch, Annelie-Saskia Wal, Maria Bader, Lennart Oswald, Franziska Faber, Luisa Lucke, Liane Otto, Christina Fenske, Martin Fritsch, Eckard Wegman, Pascal Baudner, Philipp Fröhlich and to all the other bright and kind colleagues for creating the institute's welcoming and stimulating atmosphere. Thanks for many lunches, chats at the coffee machine, for yoga and running breaks, and bike trips to retreats.

I would like to thank the Faculty VI at TU Berlin for accepting me into their PhD program. I am grateful to Christoph Roesrath, Erika Krieger-von Stephani, and Dorothe Ilskens for their help preparing the submission of this dissertation.

I am grateful to Sabine Fuss and Nicolas Koch for their supervision and support. I want to thank them for encouraging me to pursue questions about which I was excited and for their patience with my occasional habit of going down data rabbit holes. Without them, I would not have gathered the confidence to stick to some of my riskier projects and I would not have had the opportunity to learn so many new things outside of my previous expertise.

I am hugely indebted to my brilliant co-authors for the hard work they put into our collaborations and the patience they had with my reasoning, workflows, and writing. I want to thank Nicolas Koch for his kindness, his calm, his focus on solutions, and his sharp thinking. I could not have imagined a more orienting, grounding, harmonious, and productive collaboration. I want to thank Valentin Guye and Hannah Klauber for their creative and able analyses, their impressive coding skills, their tireless work, their equanimity and their boundless motivation. I want to thank Robert Heilmayr, Jacqueline Liu, and Sabine Fuss for their great ideas, their precise writing, and their benevolence in our collaborations.

I had the privilege to work with stellar research assistants, whose brilliance amazes me and who will, without any doubt, have hugely impactful careers. Thanks to Claudia Günther, Hanif Kusuma Wardhani, Tarun Khanna, Ben Thies, Lennard Naumann, and Mayang Krisnawardhani for their hard work.

I want to thank all those who allowed me to do field work or exchange with academics in other places. I am grateful to the team at SMERU in Jakarta for hosting me, to Robert Heilmayr and Jason Jon Benedict for welcoming me to their floor at the Bren School and at their seminar, and to the Restore+ team – Florian Kraxner, Ping Yowargana, Charlotte Kottusch, Constantino Dockendorff, and many others – for their feedback and for organizing several inspiring workshops.

Many other colleagues and friends have given me feedback and advice. I am very grateful that people were willing to take time out of their busy schedules and listen to the often very niche problems about which I asked them. Many colleagues gave me feedback at seminars, on Twitter or after personal, sometimes even unsolicited, email requests. I built on the work of many generous people who shared their ideas, their code, and their data. I am very grateful for this and it makes me very hopeful for the future of academia, despite the competition that external incentives tend to favor in the profession as a whole. I hope that I included all of them in the acknowledgements of the individual chapters of this dissertation. I want to particularly thank Valentin Guye for discussions on econometrics and tropical forest policy and Hannah Klauber for deconstructing fixed effects estimators together, and Luke Haywood for giving me honest feedback on any of my ideas, no matter how far-fetched. I am extremely grateful to Leonard Missbach, Hannah Klauber, and Sabine Fuss for giving me feedback on the introduction and conclusion of this dissertation. I want to thank Alex Rohlf for bearing with my basic questions and to Max Callaghan, Hauke Ward, and Jérôme Hilaire for helping me with code. I want to thank all my other collaborators – among them Barbora Sedova, Leonard Missbach, Jan Steckel, Sebastian Renner, Victoria Wenxin Xie, Gregor Schwerhoff, and Christopher Bren d'Amour – for our projects, some still ongoing.

I want to thank Johanna Wehkamp for inspiring me to write this dissertation. I am grateful for her intelligence, her relentless optimism, her humanity, and her belief in people. Thank you for your advice, for challenging and for encouraging me: thank you for making me feel safe and loved, and for making me laugh.