The Distributional Consequences of Incomplete Regulation

Danae Hernández-Cortés*

June 2021

Abstract

Environmental policies can be ineffective if firms shift production processes from regulated to unregulated sectors. Such incomplete regulations could affect the spatial distribution of pollution and who bears its burden. I study this phenomenon in the context of a policy intended to reduce pollution from industrial mills that process sugarcane in Mexico. In response to the regulation, mills shifted processing to agricultural fires on sugarcane fields. Following the policy, sugarcane fields linked to regulated facilities increased fires by 14% resulting in a 6% increase in $PM_{2.5}$ concentrations. This pollution increase is associated with decreased birth weight for children in nearby populations, I find an increase in the likelihood of very low weight by 3% and very preterm birth by 2%. Pollution increases are unevenly distributed across communities since agricultural fields tend to be located near poorer populations. These findings highlight an often undiscussed implication of incomplete pollution regulation: its distributional consequences.

*University of California, Santa Barbara, hernandezcortes@ucsb.edu. I appreciate the detailed feedback and support from my committee, Kyle Meng, Kelsey Jack, and Christopher Costello. I also want to thank Noé Aguilar Rivera for sharing his data and CONADESUCA officials, José Fernández Betanzos and Pedro Aquino Mercado, for all their support accessing the data. This project has benefited from comments by Amy Ando, Mark Buntaine, Peter Christensen, Tatyana Deryugina, Olivier Deschenes, Joel Ferguson, Matthew Fitzgerald, Don Fullerton, Robert Heilmayr, Juliana Helo, Andrew Plantinga, Paulina Oliva, Chris Severen, Eric Verhoogen, and members of the UCSB Environmental and EnvDev Reading Groups.

Introduction

Environmental policies regulate productive activities that generate pollution. However, production can be reallocated across locations and within supply chains, escaping the reach of regulation. In such circumstances, the regulation may be "incomplete" and a firm may substitute production from a regulated to an unregulated activity, generating additional pollution than absent regulation (i.e. leakage). Incomplete regulation has well established efficiency consequences (Gibson, 2018; Fowlie, 2009; Fowlie & Reguant, 2020). Much less attention has been paid to how incomplete pollution policies can create winners and losers. It is possible that individuals living near regulated activities experience relative decreases in pollution while individuals near unregulated activities experience higher pollution levels.

This paper provides evidence of supply chain leakage and its distributional consequences in the context of the sugarcane industry in Mexico. Mexico is the world's sixth largest sugarcane exporter and the sugarcane industry is an important part of the economy in southern and central Mexico. However, sugarcane harvest is a heavily polluting activity: it often requires fires in order to clean and cut the sugarcane. These fires have been found to affect in-utero health outcomes for affected populations (Rangel & Vogl, 2019).

The sugarcane sector in Mexico is an interesting setting for studying the distributional consequences of incomplete regulation. Sugar mills have two technological options for harvesting sugarcane: mechanical or manual cut. When sugarcane is cut manually, it must also be burned since fires clean excess vegetation on the sugarcane plant. If sugarcane is not cleaned in the field using fires, it needs to go through an additional cleaning process that uses industrial boilers. Starting in 2014, the Mexican government implemented a policy aimed at decreasing sulfur dioxide emissions from industrial boilers, requiring industrial facilities in all sectors of the economy using oil as fuel to reduce emissions each year by either substituting to less polluting boilers or acquiring abatement technologies. Facilities using biofuels were exempt from complying. Using rich data on sugar mills operations, technology, and production inputs and outputs, I show that regulated mills (facilities not using biofuels) shifted some of their processing to fields where sugarcane is grown, increasing the amount of agricultural fires relative to unregulated mills. Furthermore, I show that mills shifted the type of inputs used, altering the spatial distribution of pollution and the populations exposed to pollution.

The first part of the paper provides a conceptual framework to provide intuition for the expected effects under the Mexican regulation. A producer decides the optimal amount of inputs, and generates pollution associated with the use of either input. There are two types of households: low and high wealth. Households choose consumption goods, leisure, and medical expenditures to maximize utility but are affected by pollution via damages in health. A social planner implements an emissions tax to only one input, therefore the regulation being incomplete. This simple framework has three main predictions, the regulation: 1) decreases the use of the regulated input, 2) increases the use of the other input (and its emissions) conditional on inputs being gross substitutes, and 3) will be regressive, affecting low-wealth households unless medical treatment goods are relatively low cost compared to the ratio of non-labor income and wages. Therefore, incomplete regulation can create damage to low-wealth households via pollution exposure when pollution of the unregulated input increases.

The paper main contribution is an empirical analysis of the effects of incomplete regulation in Mexico's sugarcane sector. I analyze whether processing in regulated facilities (using non-biofuel boilers) was substituted with burning in the fields, as well as the resulting pollution consequences. A difference-in-differences research design compares regulated and exempt facilities before and after the policy was implemented. I find that fields linked to regulated mills increased the number of sugarcane fires by 14% following the regulation and ambient concentrations of $PM_{2.5}$ over these fields by 6%. This is corroborated using evidence of substitution responses to the regulation using detailed data on various inputs and outputs for sugarcane mills and fields. Consistent with an increase in fires, I find that fields linked to regulated mills increased manual cut workers by 5% and that the amount of sugarcane harvested using manual cut increased by 9%. I find no evidence of a change in the quantity of sugar produced as a result of the regulation and suggestive evidence of a decrease in processing efficiency.

The paper then examines whether this change in pollution disproportionately affected vulnerable rural areas. Similar to other developing countries, the agricultural fields in Mexico are located near rural areas that have higher levels of poverty and socioeconomic vulnerability. Given my findings that fires and pollution in the sugarcane fields increased, groups living near the sugarcane fields could experience an increase in pollution as a result of the regulation. I use the results from my empirical analysis to predict pollution concentrations in the fields as a result of the policy and link this to exposed populations. Populations with greater socioeconomic vulnerability experienced relatively lower pollution prior to the policy but experienced a larger share of the pollution increases due to the policy. I find that the most vulnerable households experienced the largest increases in pollution relative to less vulnerable populations. These results highlight an important finding that has not been previously empirically documented in the literature: incomplete regulation can contribute to environmental inequality by altering the spatial distribution of pollution.

Finally, this paper estimates whether the increase in pollution caused by input substitution is associated with worse health outcomes in affected areas. I use individual birth records for the period 2012-2017 obtained from the Mexican Health Ministry to estimate the impacts of pollution exposure on birth outcomes such as birth weight, gestational length, very low birth weight (<1,500 g), and very preterm birth (< 32 weeks). I link the fires location to the mother's locality of residence (rural village or city) and estimate the impact of pollution on birth outcomes. I find that an additional μg/m^3 increase in pollution is associated decreases birth weight of 1 gram, increase very low birth weight incidence by 2%, and increases very preterm birth incidence by 3%. These results add to the extensive literature examining the impacts of air pollution on health (Graff Zivin & Neidell, 2013), and in particular, the impact of pollution on birth outcomes (Currie et al., 2014). Consistent with Rangel and Vogl (2018), I find that increases in pollution caused by sugarcane fires are associated with worse birth outcomes for impacted localities. These results further document the negative impacts of incomplete regulation when producers can subsitute the inputs they use and these create higher pollution to nearby communities. Using the estimates found, I calculate that an input tax of $39 USD per ton of sugarcane harvested using manual cut (average price of sugarcane is $650 USD per ton) would internalize the impacts of sugarcane burning on affected populations.This input tax accounts for the existing regulation of boilers and its leakage to sugarcane fields as well as the health impacts of sugarcane harvesting.

This paper provides two main contributions. First, it contributes to the leakage literature by documenting a specific mechanism through which firms substitute pollution from regulated to unregulated sources: input substitution. If these substitution patterns increase emissions or generate additional social costs, environmental regulation can backfire, creating more pollution than otherwise absent regulation. Others have found that the amount of leakage induced by a regulation depends on the structure of the sector and the producers' responses to the regulation (Fowlie, 2009; Baylis et al., 2014). Firms can substitute pollution to unregulated media (Gibson, 2018), sectors (Hansman et al., 2019), other facilities (Rijal & Khanna, 2020), and countries with laxer regulations (Hanna, 2010; Ben-David et al., 2018). By focusing on one sector and using detailed individual production data, I am able to unravel how firms can alter the production processes to adjust to the regulation. This paper highlights a previously overlooked mechanism through which incomplete regulation can create leakage: firms can shift towards dirtier, unregulated inputs. In addition, this paper provides evidence of leakage from point sources (industry) to non-point sources (agricultural fires). Shifting pollution from point sources to non-point sources could be particularly problematic since non-point sources are harder to regulate due to their dispersed nature (Shortle & Horan, 2001). In so doing, I contribute to another literature that explores the role of regulation in incentivizing firms or individuals to adjust margins to avoid regulation (Carrillo et al., 2017; Yang, 2008).

Second, I contribute to the literature on environmental justice and inequality in the distribution of pollution. Agricultural fields are mainly located in rural areas, that are on average poorer and face higher socioeconomic vulnerability than their urban counterparts. By increasing the number of fires and pollution in these areas, regulation aimed at point sources with the potential to reallocate production to non-point sources could increase pollution in already disadvantaged areas. The environmental justice literature has long studied the unequal distribution of environmental hazards finding that minority and poor populations face higher pollution levels than other communities (Mohai et al., 2009; R. D. Bullard, 2008; Morello-Frosch, 2004; Banzhaf et al., 2019). Studies have found that high polluting facilities and toxic waste sites are mostly located in poor and minority communities in the United States (Pastor et al., 2001; R. Bullard & Wright, 1987). Recent literature highlights the potential of environmental regulations to have distributional consequences across sectors in the economy (Fullerton & Muehlegger, 2019) and other studies have found that gains from environmental regulation are unevenly distributed across demographic groups (Hsiang et al., 2019; Holland et al., 2019; Currie et al., 2020). In the case of Mexico, studies have found that larger polluters tend to be located near poor, marginalized populations (Chakraborti, 2018; Chakraborti & Shimshack, 2020). This paper suggests another source of environmental injustice: incomplete regulation can cause pollution leakage to vulnerable populations.[1] Given that pollution damages could be higher in low income communities due to low access to health care and defensive investments, regulations that increase agricultural burning could generate disproportionate pollution damages to rural populations.

This paper also finds that despite increasing pollution in rural areas, input relocation can have positive employment benefits. Some studies have documented the tradeoff between health and local economic outcomes in developing countries. For instance, von der Goltz and Barnwal (2019) show the wealth-health tradeoff due to mining activities in Africa, where mines increase asset wealth in nearby communities but increase anemia and stunting for young children. Rangel and Vogl (2019) show that the increase in economic activity during the sugarcane harvest is also accompanied by worsening health outcomes for newborns in Brazil. By showing that local employment for manual work increases together with increases in pollution, this paper adds to the literature examining this health-local economic outcomes tradeoff. However, other studies have shown that regulation of polluting technologies might not need be accompanied by labor losses. In the case of sugarcane production, Costa and Lima (2020) show that harvest mechanization in Brazil decreased employment in the agricultural sector but increased employment in the manufacture and services sectors.

The results of this paper extend beyond the sugarcane industry in Mexico. For instance, several other studies have found supporting evidence of the "pollution haven effect" which highlights that environmental damage might be shifted towards places with less strict regulation or unregulated places (Hanna, 2010; Heilmayr et al., 2020). Other examples of where this could occur are global chains for processed food when regulation is incomplete between two countries. My results suggest that regulators attempting to regulate one input need to be aware of firms' responses in unregulated sectors and the location of these sectors relative to disadvantaged

[1] Other studies have analyzed additional mechanisms of environmental injustices in the case of the U.S, such as incomplete information about pollution damages and hidden pollution (Hausman & Stolper, 2020) and mergers (Jacqz, 2020).

communities.

The rest of the paper proceeds as follows. Section 1 describes the sugarcane sector and boiler regulation. Section 2 provides a simple conceptual framework of input substitution when regulation is incomplete. Section 3 describes the data. Section 4 presents the empirical specification. Section 5 explains the effects of regulation-induced pollution redistribution and its distributional consequences as well as the health effects of incomplete regulation. Section 6 calculates a tax that would internalize the social costs caused by sugarcane burning. Section 7 concludes.

1 Background

1.1 Sugarcane harvest and production in Mexico

Sugarcane is the main input of sugar production, which is processed in nearly 60 mills across Mexico. The high demand for sugar in Mexico (on average 80 pounds of sugar per capita consumption per year) makes sugarcane among the 10th highest demanded crops in Mexico and Mexico is the 6th largest global sugar exporter. Sugar mills have two technological options for harvesting sugarcane: either mechanical or manual cut. When sugarcane is cut manually, it must also be burned. These fires facilitate harvest by cleaning the excess of vegetation in the sugarcane plant. If sugarcane is not cleaned in the field using fires, it would need to go through an additional cleaning process that uses machines powered by industrial boilers.

Sugarcane is first harvested in the field and then sent to process at the mill. Given that there are only 60 active mills in Mexico that process nearly 865,000 hectares of sugarcane, the harvest needs to be staggered from mid-November to late May.[2] The fields are usually located within driving distance from the mills and mills in general own the fields where they source the sugarcane.[3] This also means that mill management has a decision power over the type of harvesting technique used. Important to note, there is no quality difference between sugarcane

[2]Sugarcane needs to be processed within the same week after harvest or it might lose its chaloric content, generating less sugar. However, after sugarcane has been converted into sugar, sugar can be stored for long periods. After sugar has been produced, mills send the sugar to individual packaging facilities that would distribute them for retail.

[3]Mills also report part from their production coming from private small landowners. However, the smallholders have contracts with specific mills. The mills are responsible of providing inputs to these smallholders such as machines to harvest sugarcane in the case of mechanical cut and trucks to transport the sugarcane to the mills.

cut using controlled fires and sugarcane cut using machines. After sugarcane has been cut, the sugarcane is transported to the sugar mill where it is then processed. Sugarcane cut using machines goes through an additional process of cleaning the plant that uses equipment fueled by boilers. These boilers can either use diesel, fuel oil, biofuels, or natural gas. After the sugarcane is clean, the sugarcane goes through another process to grind the sugarcane and extract its caloric content to then crystalize and refine the sugar in the mill.

The sugar producing industry is an important part of the sugarcane regions. The Mexican Agricultural Agency estimates that sugarcane production has approximately 440,000 direct employees and 2,000,000 indirect employees.[4] Although the harvest season brings employment to these regions, it comes with a cost: according to França et al. (2012), sugarcane fires are harmful because they raise particulate matter concentrations (96% of these particles are respirable), CO and NO_x which have many adverse health consequences. Moreover, sugarcane-harvest fires have been associated with negative health outcomes to nearby communities. For instance, Rangel and Vogl (2019) found that in utero exposure to pollution from sugarcane fires reduces birth weight and gestational age at birth in Brazil.[5] Sugar processing after sugarcane has been harvested is also heavily polluting. For instance, the average sugar mill generates, on average, 2,427.65 tons of NO_x per year which makes it one of the most heavily polluting industries in the country. To put in context, the average California cement facility generates 1,364.2 tons of NO_x per year.[6]

1.2 Regulating pollution from sugar mills

In 2011, the Mexican government strengthened the maximum pollution limits of all stationary sources via the NOM-085-SEMARNAT-2011. The regulation targeted many sectors including cement production, chemical manufacturing, and general industrial activities. The regulation stated that beginning in 2014-2015 all pollution sources must decrease the emissions related to

[4]Studies have documented the importance of sugar production for local employment and development. For example, Dell and Olken (2020) show that households living within a few kilomenters of historical sugar factories have 10% higher per-capita consumption than other households living further away.

[5]Other studies have shown that exposure to smoke from fires also increases early-life mortality (Jayachandran, 2009; Pullabhotla, 2018) and affects children's human capital outcomes such as exam performance (Graff-Zivin et al., 2020). Agricultural burning can also increase deaths from respiratory problems for adults (He et al., 2020).

[6]Data on mills emissions ara available for 2017 via the National Registry of Emissions (RENE) and data on California's cement emissions are available in CARB pollution mapping tool.

the combustion process.[7] The regulation stipulated that starting in 2011, the emissions from new and existing equipment must be reported to the environmental agency in Mexico and after 2015, the new emission standards need to be attained for all the combustion sources. The regulation stated an annual reduction of 1,000 ppmv of SO_2 in 2015 relative to pre-existing levels and a reduction of additional 100 ppmv per year until 2019. In case of non-attainment at the facility level, the facilities will have to pay a fee to the environmental agency depending on the exceeding emissions. Facilities that used biofuels as the main source of energy were exempt from the policy.

In the case of the sugarcane industry, the regulated equipments were mainly used as a substitute in the cleaning process. This meant regulating the boiling of (unburnt) sugarcane for facilities that are not using biofuels in their operations. As a result of the policy, non-exempt facilities could respond by either complying with the regulation, lowering the amount produced. or by shifting technologies in the field to decrease the emissions coming from the regulated technology.

The regulation is enforced by the Mexican Environmental Protection Agency's regulator entity, PROFEPA. However, in the case of the sugarcane industry, the Mexican Agriculture Agency through the sugarcane regulator entity, the *Comité Nacional para el Desarrollo de la Caña de Azúcar*, CONADESUCA, also monitors annual compliance. At the time of the regulation, biofuels were not regulated. However, there have been some recent efforts to extend environmental regulation to cover sugar mills using biofuels. For instance, the PROY-NOM-170-SEMARNAT-2017 is expected to regulate mills using biofuels once it is approved by the Mexican government.

Sugarcane burning is not regulated in Mexico.[8] Regarding agricultural fires, the NOM-015-SEMARNAT/SAGARPA-2007 specifies a few rules for agricultural burning. For instance, farmers can only burn one plot if they do not have contiguous fires in an adjacent plot. Farmers should notify neighboring plots in case of wishing to use a controlled fire and notify local

[7] The regulation in 2011 stipulated that the beginning of the compliance period should be 2014. However, in 2012 the beginning of the compliance period was extended by one year. Therefore, the beginning of the compliance period is 2015.

[8] Other countries have started regulating sugarcane burning. For example, Brazil's sugarcane growing regions have started adopting mechanical harvesting methods in the last decades and has nearly complete adoption by 2013 (Davis, 2016).

authorities if the fire grows uncontrollably. However, sugarcane harvest fires are not regulated.

In this paper, I leverage variation induced by the regulation to compare regulated and unregulated facilities before and after the regulation was introduced.[9] The following section provides a simple model of input decision with incomplete regulation and households' welfare. The objective of the following section is to provide intuition on the possible producer responses under the new regulation.

2 Conceptual framework

This section presents a conceptual framework of incomplete pollution regulation and its welfare effects to nearby populations. The objective of this section is to explain under which conditions incomplete regulation can create input substitution. Similarly, I derive conditions under which incomplete regulation can be regressive, affecting poor households via pollution damages more than when no regulation exists.

Producer maximization problem: A producer decides between labor (l) and capital (k) that are used in the production of a homogeneous good that is competitively produced. The producer is a price taker in both the labor and capital markets (input prices are w and r, respectively, and both are competitive input markets) and faces a Constant Elasticity of Substitution (CES) production function.

$$Y = [k^\sigma + l^\sigma]^{1/\sigma}$$

The use of each technology produces pollution emissions (γ_l and γ_k) that are an increasing function of input use.

$$\gamma_k = f(k) \quad \text{and} \quad \gamma_l = f(l)$$

The representative producer chooses l and k to maximize:

[9]One concern using this specification is that non-exempt facilities could change regulation status after the policy started in order to be exempt of the policy (i.e. regulation-induced technology adoption) which changes the composition of the control and treatment groups. This is an unlikely concern in this setting. In general, the decision to invest in boilers for the facility operations/electricity generation is a long-run decision, whereas this paper focuses in the three years (short run) of the policy. Moreover, I obtained data on all of the mills' industrial investments in the recent years and electricity generation permits and only one mill invested in a new boiler in 2016 (Ingenio San Francisco Ameca that acquired a biofuel boiler), most of the studied mills' last investment was done during 1980-2000.

$$\pi = p[k^\sigma + l^\sigma]^{1/\sigma} - rk - wl \tag{1}$$

The optimal share between l and k is given by:

$$l^* = \left(\frac{w}{r}\right)^{\frac{1}{\sigma}} k^* \tag{2}$$

Optimal use of l and k is given by:

$$l^* = \bar{y}\left[\frac{w^{\frac{\sigma}{\sigma-1}}}{r^{\frac{\sigma}{\sigma-1}} + w^{\frac{\sigma}{\sigma-1}}}\right]^{1/\sigma} \quad \text{and} \quad k^* = \bar{y}\left[\frac{r^{\frac{\sigma}{\sigma-1}}}{r^{\frac{\sigma}{\sigma-1}} + w^{\frac{\sigma}{\sigma-1}}}\right]^{1/\sigma}$$

Households' maximization problem: There are two types of households: low-wealth households (L) and high-wealth (H) households. Low-wealth households' utility is a function of consumption goods x_L, leisure $(1 - l_L^s)$, where l_L^s is labor supplied, and health H_L, where $H_L = \mathcal{H} + \beta M_L(\gamma_l) - D_L(\gamma_l)$.[10] $\mathcal{H}$ is a health stock that is determined by external and genetic factors, $M_L(\gamma_l)$ denotes medical care, and $D_L(\gamma_l)$ is pollution exposure created by emissions linked to the use of l.[11] Low wealth households decide the amounts of consumption goods, the labor to supply for the production of the good explained in the producers' maximization problem, and the amount of medical care to consume taking prices (c_M and c_x) as exogenous.[12] Low-wealth households obtain wages w associated with labor supplied and other non-labor sources of income, I_L. High-wealth households' utility depends on consumption goods, x_H, and health, H_H, where $H_H = \mathcal{H} + \beta M_H(\gamma_k) - D_H(\gamma_k)$. $D_H(\gamma_k)$ are the damages caused by emissions associated with the use of k and $M_H(\gamma_k)$ is the medical care, which depends on pollution emissions associated with k. High-wealth households receive a constant share of total output produced $\phi\bar{Y}$ and non-labor sources of income, I_H.

Low wealth households' maximization problem is given by:

$$\max_{x_L, l_L, M_L} \{ \quad u_L(c_L, l_L^s, M_L) = \log(x_L) + \log(1 - l_L^s) + \log(\mathcal{H}_L + \beta M_L(\gamma_l) - D_L(\gamma_l))\}$$

[10]This health expression is similar to (Graff Zivin & Neidell, 2013), however, their health function also depends on avoidance behavior and the marginal productivity of labor also depends on pollution ($w(\gamma_l)$ for this case).

[11]β is the share of medical expenses used, which means that $0 < \beta \leq 1$

[12]I also assume that damages $D(\gamma_l)$ are an increasing function of pollution exposure γ_l and the demand for medical care also increases in γ_l.

$$\text{s.t. } c_x x_L + c_m M_L(\gamma_l) = w l_L^s + I_L$$

High wealth households' maximization problem is given by:

$$\max_{x_H, M_H} \{ \quad u_H(x_H, M_H) = \log(x_H) + \log(\mathcal{H}_H + \beta M_H(\gamma_k) - D_H(\gamma_k))\}$$

$$\text{s.t. } c_x x_H + c_m M_H(\gamma_l) = I_H + \phi \bar{Y}$$

The optimal l_L^s, x_L, and $M_L(\gamma_l)$ demanded are given by:

$$l_L^s = \frac{w(c_M + \beta) - \beta I_L + c_M(D_L(\gamma_l) - \mathcal{H})}{w(c_M + 2)}$$

$$x_L = \frac{1}{c_x}\left[\frac{w(2-\beta) + \beta I_L + c_M(\mathcal{H} - D_L(\gamma_l))}{c_M + 2}\right]$$

$$M_L(\gamma_l) = \frac{\beta(I_L + w) + 2(w + D_L(\gamma_l) - \mathcal{H})}{\beta(c_M + 2)}$$

The optimal x_H and $M_H(\gamma_k)$ demanded are given by:

$$M_H(\gamma_k) = \frac{1}{2c_M}\left[I_H + \phi\bar{Y} + \frac{1}{\beta}(D(\gamma_k) - \mathcal{H})\right]$$

$$x_H = \frac{1}{c_x}[I_H + \phi\bar{Y}]$$

The effect of incomplete regulation on technology use and households' welfare: Exogenous variation from the regulation establishing pollution limits to industrial boilers allows me to quantify the effects of incomplete regulation on input substitution between k and l empirically. Mechanical cut implies higher use of boilers, therefore this technology is capital intensive and is considered as k. Manual cut, on the other hand, is labor intensive and therefore is considered as l. Given that both technologies co-exist in the sugar production process, I model the choice between each technologies under incomplete regulation. The following paragraphs describe the predicted effects of incomplete regulation on labor, capital use, emissions, and wages.

Prediction 1: a tax on capital emissions decreases the demand for capital-intensive technology.

For this prediction, let's assume that a regulator who aims at decreasing the production of emissions coming from capital introduces a tax (τ) for emissions generated by capital such that the new producers' maximization problem becomes:

$$\pi = p[k^{\sigma} + l^{\sigma}]^{1/\sigma} - rk - wl - \tau\gamma(k) \tag{3}$$

$$\frac{\partial k^{\tau}}{\partial \tau} = -\bar{y}\left[1 + \left(\frac{w}{r+\tau\gamma_k'(k)}\right)^{\frac{\gamma}{\gamma-1}}\right]^{\frac{1-\gamma}{\gamma}} \left(\frac{w}{r+\tau\gamma_k'(k)}\right)^{\frac{\gamma}{\gamma-1}} \left(\frac{w\gamma_k'(k)}{(r+\tau\gamma_k'(k))^2}\right) < 0$$

Prediction 2: a tax on capital emissions increase (decrease) the demand for labor-intensive technology if labor and capital are substitutes (complements).

Expression (2) becomes:

$$l^{\tau} = \left(\frac{w}{r+\tau\gamma_k(k)}\right)^{\frac{1}{\sigma}} k^{\tau} \tag{4}$$

The optimal demand of l is given by:

$$l^{\tau} = \bar{y}\left[\frac{w^{\frac{\sigma}{\sigma-1}}}{(r+\tau\gamma_k'(k))^{\frac{\sigma}{\sigma-1}} + w^{\frac{\sigma}{\sigma-1}}}\right]^{1/\sigma} \tag{5}$$

Taking the derivative of (5) with respect to τ:

$$\frac{\partial l^{\tau}}{\partial \tau} = -\left(\frac{\bar{y}w^{\frac{1}{\sigma-1}}}{\sigma-1}\right)\left[\frac{(r+\tau\gamma_k'(k)))^{\frac{1}{\sigma-1}}\gamma_k'(k)}{\left[(r+\tau\gamma_k'(k))^{\frac{\sigma}{\sigma-1}} + w^{\frac{\sigma}{\sigma-1}}\right]^{\frac{1+\sigma}{\sigma}}}\right]$$

By the CES properties, gross complements ($\sigma > 1$) implies $\sigma - 1 > 0$, which means that $\frac{\partial l^{\tau}}{\partial \tau} < 0$. Conversely, gross-substitutes implies that $\frac{\partial l^{\tau}}{\partial \tau} > 0$. Therefore, regulating the emissions from capital when labor and capital are gross-substitutes means a higher use of the labor intensive technology and increased emissions from its use.

Prediction 3: low-wealth households will be worse off under incomplete regulation on k unless $\frac{I_L}{w} > \frac{c_M(2-\beta)}{\beta}$. High-wealth households will only be worse off under regulation on k if $\beta(I_R + \phi Y(\tau)) < 2D(\gamma_k)c_M$. For this derivation, see Appendix A.

The conceptual framework predicts that under incomplete regulation, the emissions from the regulated technology decrease whereas the use of the unregulated technology and its emissions increase, conditional on being substitutes in the production process. This increase in pollution is regressive, affecting low-wealth households utility unless the ratio between non-labor income sources, I_L, and labor wages, w is larger than the price of medical treatment discounted by β.[13]

These results have implications for the setting studied in this paper. Under the new regulation of sugar mills, we can expect that regulating the technology used to process sugarcane in the mill translates into higher use of its substitute: manual cut. This prediction implies a shift from the capital intensive technology towards the labor intensive technology, increasing its associated pollution emissions. The following sections examine the impact of the policy on input substitution and its consequences for pollution emissions.

3 Data

This paper uses a combination of remote sensing data and administrative data. The remote sensing data allows me to measure fires, land use, and pollution. The administrative data from sugarcane producers in Mexico allows me to document input-use responses to the regulation. Combining these data sources, I created an exhaustive dataset of weekly inputs use and outputs, and daily associated fires and pollution from 2012 to 2017. This section provides a description of the data sources and the construction of all the relevant variables.

3.1 Fires data

I obtained data on the universe of daily fires in Mexico from the Active Fire Data product based on the NASA's Visible Infrared Imaging Radiometer Suite (VIIRS). This product provides data on all fires occurrences starting in February 2012. NASA detects fires in a 375 m $\times$ 375 m grid and provides the centroid of the pixel with a fire event.[14] I restricted the fires to the months November to May to cover the sugarcane harvest season because sugar mill operations are concentrated in these months.[15]

[13]Note that this condition only holds in the case that pollution does not affect productivity, (i.e. w does not depend on γ_l).

[14]The average size of sugarcane fields in Mexico is approximately 4.7 hectares (Campos Ortiz & Oviedo Pacheco, 2015). This means that the VIIRS pixel covers approximately two average fields.

[15]CONADESUCA reports the start and end of the harvesting season. For the years in the sample, the harvest begins around the third week in November and finishes at the end of May.

3.2 Sugarcane coverage data

In order to identify the extent of sugarcane fields in Mexico, I used data from Mexico's National Committee for Sugarcane Suistainable Development (CONADESUCA). The data include confidential information of sugarcane plots in Mexico. CONADESUCA used Landsat 8 images from 2014-2015 to map the sugarcane plots in Mexico. To do so, they classified sugarcane fields using the Normalized Difference Vegetation Index (NDVI) and revalidated using Landsat data and field visits.[16] Given the confidentiality of the data, CONADESUCA linked the fires centroids from VIIRS with the sugarcane fields polygons for this project. Therefore, I am able to identify whether a particular fire event occurred inside a sugarcane polygon. I obtained a total of 23,106 sugarcane fires for the study period 2012-2017.[17] These fires are classified as sugarcane-harvest fires.

3.3 Mill characteristics

The geographic location of all sugar mills was obtained from the National Statistical Directory of Economic Units from INEGI that is based on the Economic Census 2009 performed by INEGI. The location information about the mills was then linked with detailed mill-level production data provided by the Sinfocaña system updated by CONADESUCA.[18] Figure A1 shows the geographic coverage of the sugar mills along with their exempt classification based on fuel use pre-policy. The fuel use pre-policy was obtained from CONADESUCA sustainability annual reports. A mill is considered to be part of the non-exempt group if it did not use biofuels in their production process or if it did not use biofuels for co-generate electricity for its production activities during 2010 and 2011.[19] Table A1 shows descriptive statistics for facilities using

[16] After processing the NDVI, CONADESUCA calculated the average lifetime of sugarcane to estimate the plant's maximum growth in order to correctly monitor the NDVI changes. They estimated the month with the highest sugarcane height and cross-validated with other SPOT images from Landsat. Furthermore, they performed field visits to sugarcane fields in Mexico to cross-validate the information.

[17] In order to correct for measurement error between the VIIRS resolution and the sugarcane fires provided by CONADESUCA, I also created a 50 m buffer around the fires and classified as sugarcane fires other fires in the VIIRS dataset that were not classified as a sugarcane fire but that were captured at the same date, time, and within the 50 meters of the sugarcane fires. This was done in order to account for fires that are not classified as sugarcane fires. This procedure yielded a total of nearly 200 additional sugarcane fires.

[18] Source: `https://www.siiba.conadesuca.gob.mx/infocana/`

[19] I obtained information on biofuel use for 50 mills. I obtained information of the 10 additional mills that used oil either for generating electricity or oil-fueled boilers in their production using CONADESUCA annual reports, therefore being regulated by the NOM-085-SEMARNAT-2011. I cross-validated the exempt vs. non-exempt definition using a list of compliance at the mill level provided by PROFEPA.

biofuels (exempt) and oil (non-exempt). Exempt facilities have on average lower daily fires and lower mechanical and manual sugarcane harvested. The empirical specification accounts for underlying differences in these facilities by using a difference-in-differences design.

3.4 Sugarcane and sugar production data

Detailed weekly administrative data of inputs and outputs at the mill level was obtained from CONADESUCA's Sinfocaña system. The data includes information on inputs and outputs for each mill and its associated fields. Information on inputs includes the total number of field workers, total harvested sugarcane (tons and hectares), total sugarcane cut used manual and mechanical cut, hours worked, among other information from the fields.[20] The outputs information includes raw processed sugarcane, processed sugarcane per day, total sugar produced, total sugar produced per day of operation, sugar-by products like alcohol and molasses, and indicators of sugar production efficiency. The sugar mills also provide information on energy and production efficiency as well as compliance to the NOM-085-SEMARNAT-2011 (previously known as NOM-085-ECOL-1994).[21]

I also obtained agricultural wages per day for employed workers at the municipality level. The data is based in payroll contributions to the Social Security Institute (*Instituto Mexicano del Seguro Social*, IMSS). The data contains total workers by sector, age, and gender at the district (municipality) level. IMSS covers mainly formal workers and might not be a good representation of agricultural workers in subsistance agriculture areas. However, in the case of sugarcane production, sugarcane harvest workers are among the workers with social security access. The data is reported at the municipality level, not at the mill level like the rest of the results from mills, therefore, for the analysis of wages I drop the municipalities that have more than one mill within the municipality with different regulation status (two municipalities, four sugar mills total).

[20]This information includes information on the fertilizers, the number of days of active production, pests in fields, and observed temperature and precipitation.

[21]Source: `https://www.siiba.conadesuca.gob.mx/sicostossustentabilidad/consultapublica/IndicadoresPublico.aspx?app=sustenta`

3.4.1 Linking fires and sugar mills

I linked sugar mills to their distribution fields by calculating the distance from the sugarcane fires to all existing mills and associated the fires to the closest mill. Distance to the mill is likely a good indicator on property: sugarcane needs to be processed within a week of being harvested or it can lose caloric content and produce less sugar.[22] Indeed, Figure A2 shows that most of the sugarcane is processed within 48 hours after being cut. Furthermore, sugar mills usually own the sugarcane fields that supply to them and incurr in the transportation costs from the fields using their own trucks and lend mechanical harvest technologies to smallholders.[23] To the extent that there are not consistent differences between non-exempt and exempt facilities in misassignment on the fires, measurement error linking facilities and mills is likely to downward bias my estimates.

I corroborated the link by obtaining information on a random sample of the supply fields for the mills in the state of Veracruz, Mexico and calculated the overlap of the sugarcane harvest fires definition with these fires. The sample was obtained by the Mexican government in collaboration with the Universidad Veracruzana and contains sampling points of sugarcane fields in the state of Veracruz, the largest sugarcane producer state.[24] I created a buffer of 300 and 500 meters surrounding the sampling sites and compared the mills association of these fires. I found that I correctly classified 80% of the fires based on 2009 data. Figure A3 shows an example of the geographic extent of this data and the sugarcane fires data. Figure A4 shows the classification (either matched or missassigned) of the mills in this sample, showing that most fires and mills are assigned correctly, except two mills were their misassigned fires are a larger share of total fires.[25] Section 4 shows the results of the empirical specification only considering the fires inside the buffers of the sampling sites. The results show higher magnitudes

[22] Within this week, sugarcane needs to be transported from the mill to the field, wait to weighted by mill workers, cleaned, and processed.

[23] In some cases small landholders or *Ejidos* own fields of sugarcane and eventually sell the sugar to a mill. These individual transactions are very hard to track and there is no consistent record of it. However, I performed interviews to sugar mill workers and they mentioned that this is a small percentage.

[24] The data was obtained with support from Noe Aguilar Rivera who shared the data of the project "Digitalización del Campo Cañero en México para Alcanzar la Agricultura de Precisión de la Caña de Azúcar". This was a sampling effort from part of the Mexican government and the Universidad Veracruzana to collect data of a random sample of the sugar mills and their corresponding sugarcane fields in the state of Veracruz in 2009.

[25] One limitation of this study is the lack of data on field ownership for other states. However, the distance assignment is likely a smaller problem in states were sugar mills are located farther away from each other, given the biological and production characteristics of sugarcane processing.

and significance level compared to the full set of fires.

Figure A5 shows the distance distribution of the sugarcane fires with respect to mills. As a comparison, I also show the distance to other fires, most of the sugarcane fires are within 20-70 kilometers from a sugar mill, which is consistent with field interviews to sugar mill administrative staff. Important to note, sugar mills sell the sugarcane that they harvest to other mills when they do not have enough capacity to process it within the week it was harvested. This could be a problem since I would be miscalculating the amount of sugarcane processed with either biofuels or oil/coal. Figure A6 shows the total sugarcane harvested by the own mill and the amount either sold or received from other mills. The vast majority of sugarcane processed by the mills is originally harvested either in their plots or in fields owned by small landholders that sold the sugarcane directly to the mills.

3.4.2 Other agricultural fires

To perform a falsification test, I obtained the number of agricultural fires that are not associated to sugarcane areas in order to obtain a fires "placebo" group. I used the rest of the non-sugarcane fires in the VIIRS data not classified as belonging to a sugarcane polygon by CONADESUCA. I classified them as agricultural fires (NSHF) if they were within an agricultural land pixel using 2012 land use data from the Mexican Land Use data series V (*Serie V de Uso de Suelo y Vegetación*) from INEGI. I followed the same procedure to classify each fire in the mill "catchment" area than for the sugarcane fires. Figure A5 shows the distance distribution between these agricultural fires and the mills. Compared to the sugarcane fires, these fires are located further away from the mills, which is what we would expect if these plots were not used as sugarcane fields. As another robustness check, I also classified fires as "non-harvest sugarcane fires" if these fires occurred within a sugarcane field but during the months June-October, outside the usual harvest window for sugarcane.[26]

[26]These fires are usually related to sugarcane residue burning post-harvest activity and occur after the main harvesting season.

3.5 Pollution data

I obtained daily pollution data from NASA's MERRA-2 aerosol optical depth product.[27] The daily pollution data has a 0.5 ° × 0.625 ° resolution. This reanalysis AOD product has information of SO_2 and I calculated $PM_{2.5}$ following Buchard et al. (2016). This methodology is analogous to other work that uses satellite data to measure pollution in areas that are remote and without a close pollution monitoring station (Chen et al., 2017). I linked the fires and mills coordinates to the pollution pixels and calculated the pollution associated to the mill or the fire in that pixel during the day of the event (in the case of the sugarcane fires) or the day of the production season (in the case of the mills).

One limitation of the data from MERRA-2 is the spatial resolution, especially for obtaining pollution for small areas such as the location of mills. In order to address this problem, I used data from Hammer et al. (2020) that estimates global annual surface fine particulate matter ($PM_{2.5}$) for 2012-2017. These data have a resolution of 0.01°× 0.01°. The dataset provides measurements of $PM_{2.5}$ $\mu g/m^3$ from aerosol optical depth and accounts for transport of pollutants using the GEOS-Chem chemical transport model. The data has been used in other contexts for the U.S. (Fowlie et al., 2019) and its spatial definition is desirable to annalyze detailed spatial units such as mills. The downside of the data is the temporal scale since it only provides annual estimates of particulate matter.

3.6 Birth outcomes data

Data on birth outcomes was obtained from the Mexican Health Ministry (*Secretaria de Salud*) that collects data from individual birth certificates and has information on all birth records and mother's demographic and residency information such as number of doctor visits, age, education, employment, and locality of residence. I am able to link the locality of the mother's residence to the sugarcane fire catchment areas by obtaining all the rural villages and cities located within 10 km from sugarcane fields and associating the average pollution exposure in the last pregnancy trimester given the literature findings linking pollution exposure and negative birth outcomes for the last pregnancy trimester (Currie et al., 2009; Rangel & Vogl, 2018). I

[27] In specific, we used the diurnal, time-averaged, single level assimilation, Aerosol Diagnosis V5.12.4 (M2TUNXAER).

merged average daily birth outcomes at the locality (city or rural village) to the average monthly pollution exposure in each month of the last pregnancy trimester. This approach is similar to Rangel and Vogl (2018), however, I am not able to observe some of the outcomes the authors examine such as hospitalizations or fetal and neonatal mortality.

3.7 Socioeconomic characteristics

In order to analyze the distributional consequences of incomplete regulation policies and pollution leakage, I used data from the Mexican National Marginalization index constructed by the Mexican government and used to classify the socioeconomic vulnerability of urban and rural areas. The index uses several variables to calculate the marginalization level, among them the percentage of people older than 15 without education, percentage of households without piped water, bathroom, electricity, and refrigerator, and average number of people living in a household. The index uses data at the locality[28] level and classifies the localities in five levels of marginalization: very low, low, medium, high, and very high. I also use data from 2010 census in order to calculate poverty levels at the locality level in order to analyze whether poorer communities experienced a higher increase in fires.

4 Empirical Specification

4.1 Impact of incomplete regulation on within supply chain leakage

The first objective of this paper is to estimate the impact of regulation on within-supply chain leakage and pollution redistribution. I take advantage of the introduction of the regulation in 2015 to point-sources, comparing two groups: non-exempt facilities (oil) and exempt facilities (biofuels). In order to analyze whether non-exempt facilities substitute from cleaning sugarcane in the mills using regulated boilers to cleaning in the fields using controlled fires, I use a difference-in-differences approach:

$$SHF_{idm} = \alpha + \beta_1 D_i \times \mathbf{1}[t \geq 2015] + \gamma_i + \mu_t + \rho_m + W_{id} + \epsilon_{idm} \tag{6}$$

[28] In terms of urban areas a locality is analogous to a city and in terms of rural areas is analogous to a village. A caveat with this classification is that I am not able to disentangle within-city variation in urban areas (i.e. neighborhood). However, I plan to extend this in future work.

Where SHF_{imd} is the sugarcane harvest fires in day d associated to mill i as described in section 3.4.1, D_i equals one if the sugar mill is a non-exempt facility, γ_i are mill fixed effects, μ_t are year fixed effects, ρ_m are month fixed effects to control for seasonality in harvesting activities, W_{id} are weather controls, and ϵ_{idm} are two-way clustered standard errors at the municipality and year level following Cameron, Gelbach, and Miller (2012). β_1 shows the difference-in-difference estimate of the impact of being non-exempt from the new emission limits after 2015.

The identifying assumption of equation 6 is that in the absence of treatment, fires in both exempt and non-exempt facilities would have followed the same trend. Testing this assumption is not possible but showing parallel trends in the outcomes of interest allows me to informally test for differences in the groups prior to the introduction of the policy.[29] Figure 1 shows that pre-treatment, both exempt and non-exempt facilities follow similar trends in the number of daily fires. This figure suggests that prior to the start of the program, fires in exempt and non-exempt facilities followed a similar trend. I performed two falsification tests. First, I replaced the dependent variable for $NSHF_{it}$ that denotes the number of agricultural fires in non-sugarcane plots associated to mills. Second, I restricted the timeframe of the fires to the months of June through October, outside of the harvesting season. There is no significant increase in NSHF after the policy and there is no increase in SHF outside the harvesting season. Figure A8 shows the parallel trends controlling for international sugar prices and Mexican crude oil prices.[30]

In a similar way, I examine whether the changes in the number of fires are associated with the substitution of inputs related to the fire use. Following a similar approach to equation 6, I estimate the following difference-in-differences specification:

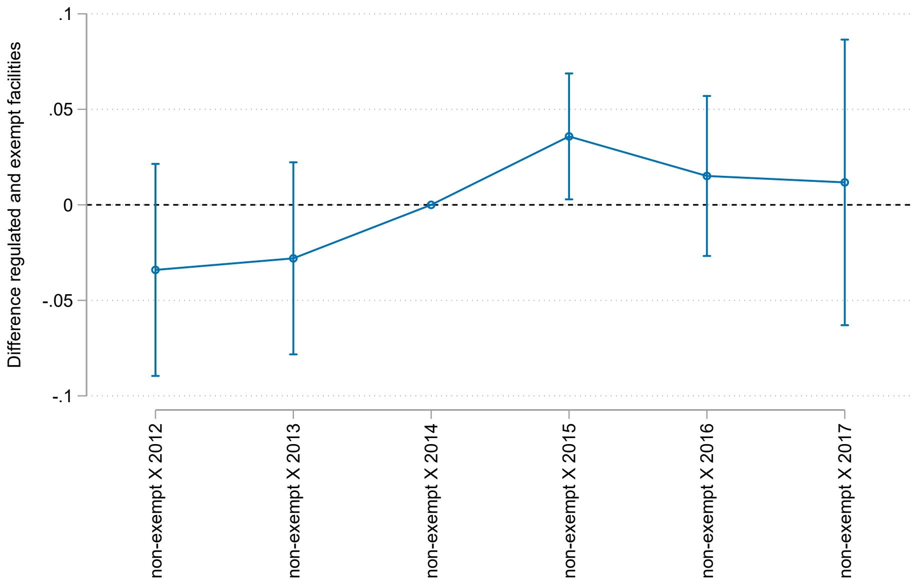
$$Y_{ist} = \alpha + \beta_1 D_i \times \mathbf{1}[t \geq 2015] + \gamma_i + \mu_t + \rho_s + \epsilon_{ist} \tag{7}$$

Where Y_{ist} denotes the variables of interest at the sugar mill level such as number of tons harvested using manual and mechanical cut, total manual workers, total sugarcane harvested, total sugarcane processed, and total sugar. s is the week with respect to the beginning of

[29]Figure A7 shows the total number of fires by exempt and non-exempt facilities. In general, regulated mills have a higher number of associated harvest fires than exempt facilities.

[30]I obtained monthly sugar prices from the Federal Reserve Economic Data (FRED) of St. Louis Fed. I used the nominal sugar prices (PSUGAISAUSDM) and the US CPI (CPALTT01USA6615) to obtain the real sugar price. I obtained daily Mexican crude oil prices from the Mexican Central Bank (SI744, *Precios del Petróleo: Mezcla Mexicana, Dolares por barril, PMI* and obtained the monthly average.)

Figure 1: Effect on total daily sugarcane fires:



Notes: This figure shows the differences in differences-year specific coefficients for the total number of daily fires following equation 6. The regulation started in 2015. 95% confidence intervals calculated using two-way fixed effects at the municipality and year level.

the harvest,[31] and ρ_s are week with respect to harvest fixed effects. Figure A9 shows parallel trends for each of the inputs: sugarcane harvested by mechanical and manual cut, total tons harvested, and number of manual workers. Figure A10 shows parallel trends for outputs such as total sugarcane processed and total sugar produced. These two figures show that there are no striking differences between regulated and exempt facilities at the start of the policy in terms of inputs used or total sugarcane harvested or sugar produced except for the amount of mechanical cut. Any differences in mechanical cut associated with the policy will not be interpreted as causal.

Finally, in order to examine whether there are differences in air pollution concentrations due to changes in the number of fires or production patterns, I use a similar specification than equation 6:

$$P_{idm} = \alpha + \beta_1 D_i \times \mathbf{1}[t \geq 2015] + \gamma_i + \mu_t + \rho_m + W_{id} + \epsilon_{idm} \tag{8}$$

Where P_{idm} is the ambient pollution concentration of $PM_{2.5}$ and SO_2 at the daily level in

[31]The administrative data is reported by sugar mills directly in a weekly basis and they start reporting it at the beginning of each harvesting cycle. However, information on the date of the beginning of the harvest for each mill is not available.

$\mu g/m^3$ obtained using the pollution level described in the Data section. I ran two separate versions of equation 8: one for the pollution associated with the pollution level in the fires polygons and another for the pollution associated with the mills' location. Figure A11 shows the parallel trends for the pollution associated with the fields and Figure A12 shows the pollution parallel trends associated with the mills.

4.2 Distributional effects of incomplete regulation

The second objective of this paper is to analyze the distributional consequences of within supply chain leakage. A large body of literature has documented negative effects of pollution on health outcomes (Rangel & Vogl, 2018; Chay & Greenstone, 2003; Deryugina et al., 2019) and how the damages of pollution could vary across income levels (Arceo et al., 2016). Other studies have analyzed whether the damages of environmental policies are distributed unevenly across populations (Fowlie et al., 2012; Grainger & Ruangmas, 2018). However, studies that document emissions leakage caused by a policy have not examined the extent of which the emissions are distributed across populations.[32] Understanding how the damages of environmental policy are distributed across populations and the determinants of the environmental damages is important for welfare analysis (Hsiang et al., 2019) and environmental justice (Banzhaf et al., 2019). Furthermore, analyzing the distributional damages of incomplete regulation in the context of this paper is important because of the characteristics of the underlying population living close to the sugarcane fields. Figure A13 shows the characteristics of the populations exposed to mills and fields. In general, poorer households tend to live in rural areas that are exposed to sugarcane fires.

In order to explore the distributional consequences of incomplete regulation, I calculated the catchment areas of all localities (either urban or rural) by creating a buffer of 10 km surrounding the centroid of the locality.[33] I then merged these catchment areas to pollution concentrations by predicting the pollution exposure coming from the policy in equation 8, obtaining the predicted

[32]Important to clarify, many studies that have documented leakage have done it in terms of GHG emissions where emissions occur is not as worrying due to the nature of GHG emissions.

[33]I assigned pollution from fires and mills by calculating a receptor catchment area of 10km from the centroid of the urban or rural locality. A caveat of this analysis is that pollution can follow non-uniform transport and dispersion patterns. In future work I aim to characterize this using prevailing wind approaches. The total people in the buffer area of fires originated in sugarcane fields is 9,834,436 and the total people in the buffer area of the mills is 5,723,850.

$PM_{2.5}$ and SO_2 from the policy, $\hat{P}_{idm}$, and modifying the empirical specification of Hernandez-Cortes and Meng (2020):

$$\hat{P}_{idm} = \gamma_0 + \gamma_1 \mathbf{1}[DAC_i] + \gamma_2 DAC_i \times \mathbf{1}[t \geq 2015] + \tau_i + \mu_t + \epsilon_{idm} \tag{9}$$

Where $\hat{P}_{idm}$ is the predicted pollution exposure coming from the policy calculated in equation 8 and DAC is an indicator variable that equals one if the locality is disadvantaged (high or very high marginalization index), τ_i are mills fixed effects, μ_t are year fixed effects. Standard errors are clustered at the locality level. $\gamma_2 > 0$ implies that disadvantaged communities have experienced a higher burden of the pollution change due to the incomplete regulation than other disadvantaged localities, $\gamma_1 + \gamma_2$ is the total effect of the pollution exposure gap between disadvantaged communities and non-disadvantaged communities. I also further divide localities into the marginalization categories to examine heterogeneity across different marginalization levels. I weighted equation 9 by population to account for differences in population in disadvantaged and non disadvantaged areas.

4.3 Health impacts of incomplete regultation

This section analyzes whether the pollution exposure caused by incomplete regulation translate into negative outcomes for the populations located within the catchment area described in the previous section. In particular, I use the predicted pollution exposure derived in equation (8) to explain changes in birth outcomes:

$$H_{jd} = \alpha + \beta_1 \hat{P}_{i(d-w)m} + X_{id} + \lambda_i + \mu_t + \epsilon_{id} \tag{10}$$

Where H_{jd} denotes average birth outcomes such as birth weight, gestation length, very low birth weight, and very preterm births at the locality and day level. $\hat{P}_{i(d-w)m}$ is the predicted exposure coming from the policy calculated in equation (8) associated with the weeks w before the birthdate in the last trimester of the pregnancy, where $w \in \{4, 8, 12\}$. X_{id} are controls such as mother age and total doctor visits averaged at the locality and day level. μ_t denotes year fixed effects. Standard errors are clustered at the locality level. This specification differs from Rangel and Vogl (2018) since the authors explore the differences between upwind and downwind fires from the mother's municipality in order to isolate the impacts of pollution from the economic activity derived from the harvesting season. To the extent that $\hat{P}_{i(d-w)m}$ is obtained using variation that exploits the introduction of regulation to sugarcane mills with a rich set of controls and fixed effects, my specification is likely capturing pollution and not economic activity.[34]

5 Results

5.1 Effects on upstream fires

This section discusses the effects of incomplete regulation on within supply chain leakage. Given that fires are a production substitute for cleaning in the plant, we would expect the amount of fires to increase after the boiler regulation for regulated facilities (non-biofuel users). Column (1) of Table 1 shows the difference-in-differences estimator, β_2, of interest. This shows that there is an approximate 14% increase in the number of daily fires after the policy began. Column (2) shows the impact on the number of fires using a Poisson model and the results are similar to column (1), the increase in the count of fires is around 13.5% after the policy began.

Table 1: Effect of emission limits on daily fires

	(1) Total SHFs	(2) Total SHFs
After 2015 × non-exempt	0.04120**	0.13506**
	(0.01429)	(0.06518)
Pre 2015 mean	0.286	0.292
Obs.	71,535	70,280
R-squared	0.091	
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mun-year	Robust
Poisson	No	Yes

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of fires using equation 6. Column (2) shows the same specification but the dependent variable is an indicator variable on whether there is a fire or not in that field. Column (3) estimates the same specification in equation 6 using a Poisson model with robust standard errors. Standard errors for column (1) and (2) using two way clusters (munincipality and year) in parenthesis.

As robustness tests, Table A2 shows the results using the monthly number of fires at the mill level. The effect is similar in magnitude considering the number of monthly fires. These

[34]Exploring the upwind and downwind specification like Rangel and Vogl (2018) is one of the priorities for future work.

results are robust to controlling for oil and sugar prices: Table A3 shows these results. Table A4 shows the results from equation 6, analogous to column (1) of Table 1 with bootstrap standard errors. Table A5 shows the results for the first falsification test: other agricultural fires not related to sugarcane. There is no difference in the number of non-sugarcane agricultural daily fires after the policy for the non-exempt facilities compared to the exempt facilities. I conducted another falsification test where I restricted the sample to the fires outside the sugarcane harvest season. Table A6 shows these results, showing no effect in the increase of fires outside the harvest season. Table A7 shows a stronger effect when restricting the dataset to the fires inside of the mills' distribution areas using the sampling points of several mills in the state of Veracruz. Results in Table A7 show that the link using distance to the mills seem appropriate: most mills usually harvest their sugarcane from the nearest fields. As a robustness test, Table A8 shows the results for the mills known to be under compliance by CONADESUCA.[35] The results are stronger when restricting the data to the compliant facilities which is what we would expect if mills are actually substituting for more manual cut. Table A9 shows the results doing two additional sample selections. Sample restriction 1 estimates the results without the only mill that invested in a biofuel-powered boiler in 2016 who could have changed fuel use as a response to the policy. Sample restriction 2 estimates the results without the mill that shows a higher rate of mismatched fields based on the minimum distance definition according to Figure A4.

5.2 Effects on input substitution

Next, I turn to analyze whether the change in the number of fires is reflected in input substitution across firms. Consistent with the finding of an increase in the number of fires used during the harvest, Column (1) of Table 2 shows that there is an increase of 9% in the total sugarcane harvested using manual cut. In the opposite direction, I find that the amount of sugarcane harvested decreases although this result is not statistically significant. Given that the use of fires is consistent with an increase on manual cut, I also find that the number of field workers increase by 5%, as column (3) of Table 2 shows. The results of table 1 and 2 show that

[35] CONADESUCA verifies compliance by doing inspections to the mills every two years. Important to mention, the regulation is enforced by the Mexican Environmental Protection Agency, PROFEPA. However, CONADESUCA verifies mills' inventory every two years and reports its compliance in order to build the Sugarcane Sustainability Index.

incomplete regultation generates within supply chain leakage and changes in the inputs used. Using data on payroll for formal sugarcane agricultural workers at the municipality level, Table A10 shows that there is no change in wages for agricultural workers throughout the period of study, even dividing by different age categories. I cannot find a discernable impact on wages with the data available.[36]

Table 2: Effects on inputs substitution

	(1) Manual cut (tons)	(2) Mechanical cut (tons)	(3) Total field workers
After 2015 × non-exempt	2,913.460**	-274.228	81.893**
	(1,188.58879)	(814.47535)	(34.54865)
Mean	30,971.456	6,915.284	1,453.505
Obs.	6,095	6,102	5,640
R-squared	0.749	0.727	0.913
Year FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Cluster level	Mun and year	Mun and year	Mun and year

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the amount of sugarcane harvested using manual cut (tons) following specification 7. Column (2) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the amount of sugarcane harvested using mechanical cut (tons) following specification 7. Column (3) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of manual labor workers following specification 7. Standard errors using two way clusters (munincipality and year) in parenthesis.

Although there is an increase in manual cut, I do not find evidence of an increase in total sugar produced. Table A11 shows that there is no increase in the total amount of sugarcane processed in the mill (column 1) and no increase in the total amount of sugar produced in the mill (column 2). These results suggest that there is redistribution between input use but this does not imply that final output increased.[37] I find suggestive evidence that the non-increase in total sugar produced is due to changes in production efficiency. I estimate that changes in production efficiency by using three indicators such as (1) the total kilograms of sugar obtained by ton of harvested sugarcane, (2) total kilograms of sugar obtained by ton of processed sugarcane, and (3) sugar extraction efficiency. Table A12 show these results which

[36] Other papers have investigated whether proximity to industrial facilities have positive employment effects despite the negative pollution exposure impacts on nearby communities. Ash and Boyce (2018) found that the share of pollution risk accruing to minority groups located near polluting facilities exceeds their share of employment and wages.

[37] A potential problem conflicting event influencing the total amount of produced sugar is the soda tax that started in 2014. This tax was a flat rate per liter of soda and it was uniform across the country. However, it is not clear why this would affect my main identification strategy. (1) There is no a priori reason why a soda tax would have affected biofuel facilities differently from non-biofuel facilities. (2) Moreover, the Mexican sugar tax was levied in the consumers directly and thus any effect would be driven by sugar demand not total production.

suggest that there is a non-significant decrease in overall sugar production efficiency, consistent with the fact of non-positive effects in production due to the change in inputs.

5.3 Effects on ambient pollution

What does supply chain leakage mean in terms of total pollution generated? Table 3 shows the implications of an increase of fires in terms of local ambient pollution level around the sugarcane fields (columns 1 and 2) and mills (columns 3 and 4). Columns (1) and (2) and columns (3) and (4) are estimated using different datasets, given the spatial resolution of the data: columns (1) and (2) are estimated using data from MERRA 2.0 with calculations following Buchard et al. (2016) and columns (3) and (4) are estimated using Hammer et al. (2020). I find that there is an increase of 1.05 $\mu g/m^3$ of $PM_{2.5}$ or a 6% increase in pollution coming from the fields associated to the 2015 regulation. I do not find a change in the SO_2 levels in the fields region. Columns (3) and (4) of Table 3 show the results of pollution from the mills, suggesting a decrease of pollution near the mills following the introduction of the new regulation. These results translate into a decrease of 3% on $PM_{2.5}$ pollution coming from mills.

Table 3: Effect on pollution

	(1)	(2)	(3)	(4)
	Pollution in fields		Pollution in mills	
	PM2.5	Log(PM2.5)	PM2.5	Log(PM2.5)
After 2015 × non-exempt	1.05187**	0.04638*	-0.38421**	-0.05121
	(0.37124)	(0.01972)	(0.09956)	(0.02569)
Mean	17.549	2.608	12.566	2.455
Obs.	20,489	20,489	295	295
R-squared	0.466	0.566	0.827	0.967
Year FE	Yes	Yes	No	No
Month FE	Yes	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	No
Cluster level	Mun&year	Mun&year	Mun&year	Mun&year

Notes: Columns (1)-(4) shows changes in pollution concentrations in the fields associated to mills. Columns (5)-(6) shows changes in pollution concentrations in the mills, data of SO_2 at the small-scale mill resolution is not available. Column (1) and (2) show the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the ambient pollution level of $PM_{2.5}$ and $\log(PM_{2.5})$, respectively following specification 8 using Buchard et al. (2016). Column (3) and (4) show the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the ambient pollution level of SO_2 and $\log(SO_2)$, respectively following specification 8 using Buchard et al. (2016). Column (5) and (6) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the ambient pollution level of $PM_{2.5}$ following Hammer et al. (2020) (SO_2 data for this data product is not available). Standard errors using two way clusters (munincipality and year) in parenthesis.

The magnitudes of my results are consistent with existing studies. I find that as a response to the regulation, regulated mills increased fires by 14%. Gibson (2018) finds that regulated facilities under the Clean Air Act increase their production in unregulated facilities by 11% and Hanna (2010) finds that regulated facilities under the Clean Air Act increased foreign output by 9%. In terms of increase in pollution, I find that incomplete regulation increased pollution exposure by 6% in rural areas located near the agricultural fields. This increase in pollution is higher than the documented by Garg et al. (2020), who find an increase of 1.25% in $PM_{2.5}$ exposure as a result of increasing agricultural fires due to labor exits in India.

5.4 Who experiences the increased pollution?

This section analyzes whether vulnerable communities experienced a larger increase in pollution coming from sugarcane fires after the policy. I classified vulnerable communities using the marginalization index provided by the Mexican government. In this section I will use the official index that classifies communities from "very low" to "very high" marginalization and a classification of "disadvantaged" if the community has "very high" and "high" marginalization levels. Given the spatial distribution of vulnerable communities in Mexico, we could expect that there are differences in baseline characteristics between disadvantaged communities and non-disadvantaged localities. These differences are captured by γ_1 in equation 9. Table 4 shows that before the policy, disadvantaged localities have lower levels of $PM_{2.5}$ and slightly insignificant, higher levels of SO_2 levels. However, after the policy the total pollution exposure in disadvantaged localities increased, meaning that total pollution exposure in disadvantaged localities increased after the policy compared to non-disadvantaged localities. Exposure to SO_2 in the other hand, decreased but the difference is not significant. I estimate that by the end of the policy, disadvantaged localities experienced 12% more pollution exposure than non-disadvantaged localities relative to the no policy counterfactual.

These pollution redistribution impacts are heterogenous with respect to different levels of marginalization: the highest level of marginalization has a higher burden of pollution exposure than localities with low or very low marginalization indices. Panel a) supbanel (a) of Figure 2 shows these results. Panel a) subpanel (b) shows a robustness check using another index of social vulnerability, the "Social Lag Index", calculated by CONEVAL.[38] Figure 2 shows

[38]This index considers data from the 2010 data based on different variables than the Marginalization Index calculated by CONAPO. The index considers indicators of infrastructure at the locality level and asset holding characteristics for the localities' households.

Table 4: Distribution of the effects of pollution coming from fires

	(1) Predicted $PM_{2.5}$ (fires)
DAC=1	-0.13900***
	(0.04193)
DAC=1 × After 2015=1	0.27055***
	(0.08194)
Obs.	3,627,238
R-squared	0.717
Year FE	Yes
Month FE	Yes
Mill FE	Yes
Cluster level	Municipality
Pop. weighting	Yes

Notes: Columns (1) and (2) show the predicted difference in pollution exposure for disadvantaged localities before the policy (DAC=1) and after the policy (DAC=1 × After 2015=1. Localities exposure was calculated using catchment areas: 10km circle surrounding the locality sector. Disadvantaged localities were classified using the 2010 Marginalization Index calculated by CONAPO. Regression uses population weights using 2010 Census data.

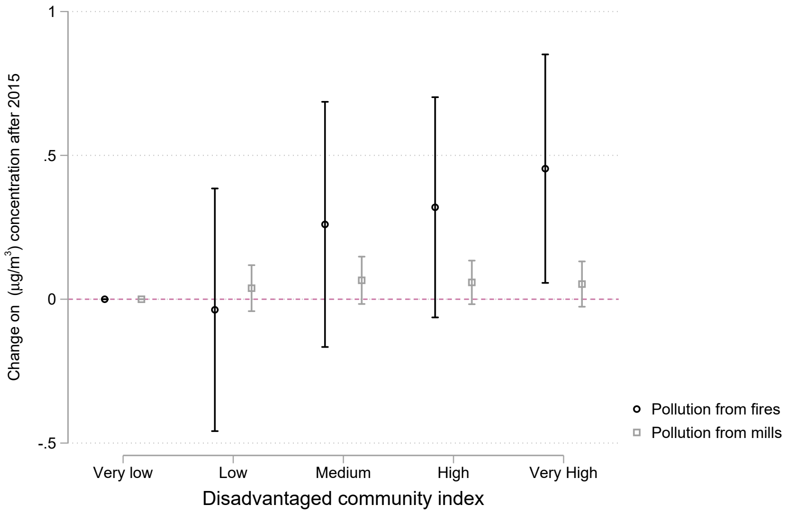
that the exposure from mills did not change differentially for communities with higher levels of vulnerability, except for the "Very High" category of the social lag index. However, it is smaller than the magnitudes from fires. The heterogeneity in pollution exposure after the policy could be explained by different reasons such as ex-ante vulnerability or because mills with fields closer to disadvantaged communities could strategically pollute more near these areas without facing opposition to pollute in these areas. Chakraborti and Shimshack (2020) show that in the case of Mexico, disparities could be explained by community pressure and collective action responses. However, additional work can be done trying to analyze whether this is the case in the sugarcane sector. These results are relevant given that previous research shows that pollution damages are not linear with income (Arceo et al., 2016). Moreover, given that deffensive investments are an important part of the willingness to pay for pollution reduction (Deschênes et al., 2017) and they could be correlated with income, poorer households might not be able to cope with changes in pollution exposure.

Table 5 shows the same specification in equaltion 9 for the mills (SO_2 is not available for the small scale resolution). There is no significant difference between disadvantaged communities and other communities after the policy began. Similarly, panel b) of figure 2 show no significant difference for the different levels of marginalization.

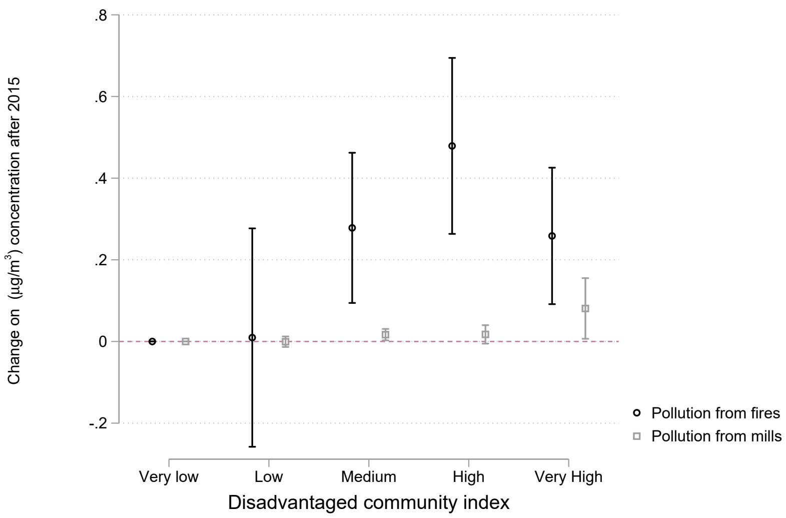
Besides showing that the policy generated a relatively higher pollution exposure for the most

Figure 2: Distribution of the effects of pollution

Panel a) Social Vulnerability Index



Panel b) Social Lag Index



Notes: Panel a) shows the results of pollution associated to fires by level of marginalization according to the marginalization index and the social lag index. Panel b) shows the results of pollution associated to mills by level of marginalization according to the marginalization index and the social lag index. Coefficients show the interaction between the marginalization level and an indicator after the policy. Regressions include mill fixed and year fixed effects. Confidence intervals calculated using clustered standard errors at the locality level. Confidence intervals calculated using two-way fixed effects at the municipality and year level.

vulnerable, I tested whether locations with higher poverty levels are the most affected by the fires. Table A13 shows that the increase in fires is higher in localities that have poverty levels higher than each locality's state median. This is consistent with Chakraborti and Shimshack (2020) who found that pollution releases is higher in marginalized communities in the case of Mexico. This is another indicator that the most vulnerable communities were affected by incomplete regulation.

Table 5: Distribution of the effects of pollution coming from mills

	(1) Predicted $PM_{2.5}$ (mills)
DAC=1	-0.01557
	(0.01044)
DAC=1 × After 2015=1	0.03893
	(0.02610)
Obs.	17,075
R-squared	0.998
Year FE	Yes
Month FE	Yes
Mill FE	Yes
Cluster level	Municipality
Controls pop.	Yes

Notes: Columns (1) and (2) show the predicted difference in pollution exposure for disadvantaged localities before the policy (DAC=1) and after the policy (DAC=1 × After 2015=1. Localities exposure was calculated using catchment areas: 10km circle surrounding the locality sector. Disadvantaged localities were classified using the 2010 Marginalization Index calculated by CONAPO. Regression uses population weights using 2010 Census data.

5.5 The impacts of incomplete regulation on health outcomes

Do changes in pollution caused by incomplete regulation affect health outcomes? I examine this by analyzing whether the predicted pollution obtained in section 4.2 changes birth weight, gestational weight, very low birth weight incidence, and very preterm birth for populations located in the fires catchment area. Table 6 shows the main health results of pollution exposure in the last pregnancy trimester on birth outcomes. I find that pollution exposure in the last trimester of pregnancy significantly lowers birth weight and increases the incidence of very low birth, and very preterm birth. These effects are larger in the weeks 5 through 12 before giving birth.

The results in Table 6 imply that an additional μg/m^3 of $PM_{2.5}$ in the weeks 1-8 of the last trimester of pregnancy associate with a birth weight decrease of 1 gram on average for all mothers. Rangel and Vogl (2018) estimate that a unit increase of PM_{10} (in μg/m^3) caused by sugarcane fires decreases birth weight by 5.2 grams. Other estimates on the impacts of pollution oh birth weight find that a unit increase in PM_{10} exposure during the last trimester is associated with a 0.4 gram decrease in birth weight (Currie et al., 2009; Rangel & Vogl, 2018). My estimates are smaller in magnitude than Rangel and Vogl (2018) which could be due to existing differences in fire activity intensity in Brazil, as well as differences in the studied pollutants and identification strategies.[39]

[39] Rangel and Vogl (2018) estimate this by comparing upwind and downwind fires which likely provide a more

Table 6: Effect of incomplete regulation on birth outcomes

	(1) Birth weight (grams)	(2) Gestation length (weeks)	(3) Very low birth weight (weight<1,500g)	(4) Very preterm birth (weeks<32)
			Panel a)	
$PM_{2.5}$ ($\mu g/m^3$) in w-4 weeks	-0.32690	-0.00144	0.00008**	0.00014***
	[0.26571]	[0.00098]	[0.00004]	[0.00005]
			Panel b)	
$PM_{2.5}$ ($\mu g/m^3$) in w-8 weeks	-1.00148***	-0.00221*	0.00012**	0.00014**
	[0.33077]	[0.00116]	[0.00005]	[0.00006]
			Panel c)	
$PM_{2.5}$ ($\mu g/m^3$) in w-12 weeks	-0.90981**	-0.00199	0.00007	0.00006
	[0.43649]	[0.00164]	[0.00006]	[0.00008]
Observations	62,129	63,803	64,026	64,026
Mean Dep. Variable	3,217.232	38.952	0.004	0.007
R-squared	0.193	0.185	0.174	0.179
Year FE	Yes	Yes	Yes	Yes
Locality FE	Yes	Yes	Yes	Yes
Cluster level	Locality	Locality	Locality	Locality

Notes: Columns (1)-(4) show changes in birth outcomes associated with pollution exposure within 4 (Panel a), 8 (Panel b) or 12 weeks (Panel c) from birthdate. Each estimate is obtained from separate regressions. All regressions control for average mothers' age and total of doctor visits during pregnancy and year and locality fixed effects. Standard errors clustered at the locality level in parenthesis.

For the other variables analyzed, I find that a unit increase of $PM_{2.5}$ is associated with a 3% increase in the probability of very low birth weight ($< 1{,}500$g) and a 2% increase in the probability of very preterm birth (< 32 weeks). These results are consistent with other studies finding increases in very low birth weight and very preterm birth associated with an increase in fires (Rangel & Vogl, 2018).[40] These results suggest that increases in pollution due to input substitution are associated with worse health outcomes for the populations located in the fires catchment area. These estimates will be used in Section 6 to calculate a tax that would internalize the health costs of sugarcane burning originated by the incomplete regulation.

6 Internalizing the costs of burning

The previous sections showed that incomplete regulation increases the emissions of the unregulated input which impacts populations located nearby. One alternative to alleviate these damages is to regulate the associated harvest burning driven by incomplete regulation via a tax and compensate the damages to affected populations. In many settings, a solution to the

precise estimate of the impacts of pollution exposure. This approach is outside the scope of the paper but will be a priority for future work.

[40]Rangel and Vogl (2018) find that an additional z-score of fire activity per week in the last trimester of pregnancy increases the incidence of very low birth weight by 22 per 1000 and an increase in the incidence of preterm birth of 23 per 1000, although the later results are not significant.

leakage problem is to extend the Pigouvian tax beyond the jurisdiction of the regulation via a border adjustment (Fowlie et al., 2021). I consider such a policy by calculating an input tax to manual cut sugarcane levied at the mills that internalizes the existing health costs associated with agricultural burning. To do so, I obtain the difference between the social cost per unit of harvested sugarcane and the private cost of production paid by consumers and find the tax that would compensate communities affected by sugarcane burning. I then calculate this tax in the context of the sugarcane sector in Mexico using the estimated responses to the incomplete regulation.

Assuming a regulator that tries to maximize consumer surplus net of private costs and environmnetal damages, The regulator's welfare maximization problem is given by:

$$W = \underbrace{\int_0^y f(y)dy}_{\text{Consumer surplus}} - \underbrace{c(l,k)}_{\text{Prod. Cost}} \underbrace{-D(\gamma_k) - D(\gamma_l)}_{\text{Pollution damages}}$$

Where $f(y)$ is the demand for sugarcane, $c(l,k)$ is the producer cost, $D(\gamma_k)$ is the health damages from pollution emissions of the capital intensive input γ_k, and $D(\gamma_l)$ is the health damages from pollution emissions of the labor intensive input.

The firms' profit function is given by:

$$\pi = py - rk - wl - \gamma_l t$$

Where p is the price of the output y, sugarcane, l is labor, k is capital, r and w are the prices of capital and labor, and t is the tax for each unit of processed sugar. The tax that would equate the social cost per unit of sugar and the private cost of production is be given by:

$$t = \frac{w\frac{\partial L}{\partial t} + \frac{\partial D(\gamma_k)}{\partial \gamma_k}\frac{\partial \gamma_k}{\partial k}\frac{\partial k}{\partial t} + \frac{\partial D(\gamma_l)}{\partial \gamma_l}\frac{\partial \gamma_l}{\partial l}\frac{\partial l}{\partial t} + r\frac{\partial k}{\partial t} - P\frac{\partial y}{\partial t}}{\frac{\partial \gamma_l}{\partial l}\frac{\partial l}{\partial t}}$$

Using the parameters obtained in the previous sections and in existing studies summarized in Table 7, I calculate that the wedge between the social cost per unit of sugarcane harvested and the private cost of production paid by consumers is given by \$39 USD (\$789 MXN) per ton of sugarcane, which average price in 2017 was \$650 USD, meaning is a 6% tax per ton of

Table 7: Parameters for tax calculation

w	Average ag. wage	\$109 MXN
$\frac{\partial L}{\partial t}$	Table 2, column (1)	9.42%
$\frac{\partial D(\sigma_k)}{\partial \sigma_k}$	Arceo et al. (2016)	0.40%
$\frac{\partial \sigma_k}{\partial k}\frac{\partial k}{\partial t}$	Table 3, column (3)	-3.05%
$\frac{\partial D(\sigma_l)}{\partial \sigma_l}$	Table 6, column (3)	2%
$\frac{\partial \sigma_l}{\partial l}\frac{\partial l}{\partial t}$	Table 3, column (1)	5.98 %
r	EPA risk free rate	0.04
$\frac{\partial k}{\partial t}$	Table A12, column (3)	-49%
P	Sugarcane price per ton	\$13,100 MXN
$\frac{\partial y}{\partial t}$	Table A11, column (1)	.19%

sugarcane.

7 Conclusion

This paper showed the distributional consequences of incomplete regulation when facilities are able to shift where production occurs. I investigated the distributional consequences of incomplete regulation in the context of the sugarcane production in Mexico. By leveraging data on fires, pollution, and detailed production information on mills, this paper is able to identify the responses of producers to incomplete regulation. I found that following the introduction of regulation, regulated facilities increased the sugarcane harvest fires in their associated fields. I also found that mills adjusted input used as a result of the regulation. Consistent with an increase in fires, I showed that regulated facilities increased employment of manual workers and sugarcane harvested using manual cut. I find that the increase in fires is accompanied by an increase in pollution near fields and a decrease of pollution near mills. Finally, I find that the increase in pollution caused by the policy is associated with worse birth outcomes for exposed populations.

The responses to incomplete regulation regarding pollution are concerning due to the differences of poverty levels of the populations close to the mills and the fields. I analyzed whether the burden of pollution caused by the policy is higher for disadvantaged populations. This paper found that the pollution increase was higher for disadvantaged communities. This result contributes to the current discussions on the determinants of environmental justice. I contribute to this literature by looking a previously overlooked mechanism: incomplete regulation.

The results of the paper are relevant for current policy debates in Mexico on whether to regulate agricultural burning from sugarcane. Moreover, in 2017 the Mexican environmental agency proposed ammendments to the existing regulation to include facilities that use biofuels as main fuel (PROY-NOM-170-SEMARNAT-2017). The results of this paper show that if facilities are able to substitute production processes with fires, incomplete regulation might backfire. Therefore, considering these possible adjustment margins is important.

There are several limitations to this study. First, the pollution level estimates should be interpreted with caution given the geographic extent of pollution measures. However, I present consistent evidence that manual cut increased together with fires, which suggest that populations located near fields were exposed to pollution. To the extent that burning biomass increases pollution levels, which has been shown by other studies, populations are likely to experience higher pollution. Second, despite the efforts to link mills to their respective fields, the possibility of misassignment of this link remains. However, given the evidence using actual distribution areas for one of the states in Mexico, it is likely that this measurement error downward bias my results, which means that the effects found in the paper could be a lower bound of the real effect. In fact, I showed how when restricting fires to the recorded distribution areas, the estimates are higher. Finally, there is still need to characterize other mechanisms driving distributional concerns of environmental policy and other environmental justice implications. However, by documenting a previously overlooked mechanism, this paper contributes to the literature on disparities in environmental impacts and its implications for environmental justice.

References

Arceo, E., Hanna, R., & Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *The Economic Journal*, *126*(591), 257–280.

Ash, M., & Boyce, J. K. (2018). Racial disparities in pollution exposure and employment at us industrial facilities. *Proceedings of the National Academy of Sciences*, *115*(42), 10636–10641.

Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives*, *33*(1), 185–208.

Baylis, K., Fullerton, D., & Karney, D. H. (2014). Negative leakage. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 51–73.

Ben-David, I., Kleimeier, S., & Viehs, M. (2018). *Exporting pollution* (Tech. Rep.). National Bureau of Economic Research.

Buchard, V., da Silva, A., Randles, C., Colarco, P., Ferrare, R., Hair, J., ... Winker, D. (2016). Evaluation of the surface PM2. 5 in Version 1 of the NASA MERRA aerosol reanalysis over the United States. *Atmospheric environment*, *125*, 100–111.

Bullard, R., & Wright, B. H. (1987). Environmentalism and the politics of equity: emergent trends in the black community. *Mid-American Review of Sociology*, *12*(2), 21–37.

Bullard, R. D. (2008). *Dumping in dixie: Race, class, and environmental quality.* Avalon Publishing-(Westview Press).

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2012). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*.

Campos Ortiz, F., & Oviedo Pacheco, M. (2015). Extensión de los predios agrícolas y productividad. El caso del campo cañero en México. *El trimestre económico*, *82*(325), 147–181.

Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, *9*(2), 144–64.

Chakraborti, L. (2018). Environmental justice and toxic releases in urban Mexico.

Chakraborti, L., & Shimshack, J. P. (2020). Environmental disparities in the global south:

Evidence from toxic water pollution in urban Mexico.

Chay, K. Y., & Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics*, *118*(3), 1121–1167.

Chen, S., Oliva, P., & Zhang, P. (2017). *The effect of air pollution on migration: evidence from China* (Tech. Rep.). National Bureau of Economic Research.

Currie, J., Neidell, M., & Schmieder, J. F. (2009). Air pollution and infant health: Lessons from New Jersey. *Journal of health economics*, *28*(3), 688–703.

Currie, J., Voorheis, J., & Walker, R. (2020). *What caused racial disparities in particulate exposure to fall? New evidence from the Clean Air Act and satellite-based measures of air quality* (Tech. Rep.). National Bureau of Economic Research.

Currie, J., Zivin, J. G., Mullins, J., & Neidell, M. (2014). What do we know about short-and long-term effects of early-life exposure to pollution? *Annu. Rev. Resour. Econ.*, *6*(1), 217–247.

Dell, M., & Olken, B. A. (2020). The development effects of the extractive colonial economy: The dutch cultivation system in java. *The Review of Economic Studies*, *87*(1), 164–203.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., & Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, *109*(12), 4178–4219.

Deschênes, O., Greenstone, M., & Shapiro, J. S. (2017). Defensive investments and the demand for air quality: Evidence from the NOx budget program. *American Economic Review*, *107*(10), 2958–89.

Fowlie, M. (2009). Incomplete environmental regulation, imperfect competition, and emissions leakage. *American Economic Journal: Economic Policy*, *1*(2), 72–112.

Fowlie, M., Holland, S. P., & Mansur, E. T. (2012). What do emissions markets deliver and to whom? Evidence from Southern California's NOx trading program. *American Economic Review*, *102*(2), 965–93.

Fowlie, M., Petersen, C., & Reguant, M. (2021). Border carbon adjustments when carbon intensity varies across producers: Evidence from california. In *Aea papers and proceedings*

(Vol. 111, pp. 401–05).

Fowlie, M., & Reguant, M. (2020). Mitigating emissions leakage in incomplete carbon markets.

Fowlie, M., Rubin, E., & Walker, R. (2019). Bringing satellite-based air quality estimates down to earth. In *Aea papers and proceedings* (Vol. 109, pp. 283–88).

França, D. d. A., Longo, K. M., Neto, T. G. S., Santos, J. C., Freitas, S. R., Rudorff, B. F., ... Carvalho, J. A. (2012). Pre-harvest sugarcane burning: Determination of emission factors through laboratory measurements. *Atmosphere*, *3*(1), 164–180.

Fullerton, D., & Muehlegger, E. (2019). Who bears the economic burdens of environmental regulations? *Review of Environmental Economics and Policy*, *13*(1), 62–82.

Gibson, M. (2018). Regulation-induced pollution substitution. *Review of Economics and Statistics*(0).

Graff-Zivin, J., Liu, T., Song, Y., Tang, Q., & Zhang, P. (2020). The unintended impacts of agricultural fires: Human capital in China. *Journal of Development Economics*, *147*, 102560.

Graff Zivin, J., & Neidell, M. (2013). Environment, health, and human capital. *Journal of Economic Literature*, *51*(3), 689–730.

Grainger, C., & Ruangmas, T. (2018). Who wins from emissions trading? Evidence from California. *Environmental and resource economics*, *71*(3), 703–727.

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., ... others (2020). Global estimates and long-term trends of fine particulate matter concentrations (1998-2018). *Environmental Science & Technology*.

Hanna, R. (2010). Us environmental regulation and fdi: Evidence from a panel of US-based multinational firms. *American Economic Journal: Applied Economics*, *2*(3), 158–89.

Hansman, C., Hjort, J., & León, G. (2019). Interlinked firms and the consequences of piecemeal regulation. *Journal of the European Economic Association*, *17*(3), 876–916.

Hausman, C., & Stolper, S. (2020). *Inequality, information failures, and air pollution* (Tech. Rep.). National Bureau of Economic Research.

He, G., Liu, T., & Zhou, M. (2020). Straw burning, PM2. 5, and death: Evidence from China. *Journal of Development Economics*, 102468.

Heilmayr, R., Carlson, K. M., & Benedict, J. J. (2020). Deforestation spillovers from oil palm sustainability certification. *Environmental Research Letters*, *15*(7), 075002.

Hernandez-Cortes, D., & Meng, K. C. (2020). *Do environmental markets cause environmental injustice?* (Tech. Rep.). mimeo.

Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2019). Distributional effects of air pollution from electric vehicle adoption. *Journal of the Association of Environmental and Resource Economists*, *6*(S1), S65–S94.

Hsiang, S., Oliva, P., & Walker, R. (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy*, *13*(1), 83–103.

Jayachandran, S. (2009). Air quality and early-life mortality evidence from Indonesia's wildfires. *Journal of Human resources*, *44*(4), 916–954.

Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental justice. *Annual review of environment and resources*, *34*, 405–430.

Morello-Frosch, R. (2004). Separate but unequal? residential segregation and air quality in us metropolitan areas. *Epidemiology*, *15*(4), S134.

Pastor, M., Sadd, J., & Hipp, J. (2001). Which came first? Toxic facilities, minority move-in, and environmental justice. *Journal of urban affairs*, *23*(1), 1–21.

Pullabhotla, H. (2018). Fires, wind, and smoke: Air pollution and infant mortality. *Job Market Paper. Available here*.

Rangel, M. A., & Vogl, T. S. (2018). Agricultural fires and health at birth. *Review of Economics and Statistics*(0).

Rangel, M. A., & Vogl, T. S. (2019). Agricultural fires and health at birth. *Review of Economics and Statistics*, *101*(4), 616–630.

Rijal, B., & Khanna, N. (2020). High priority violations and intra-firm pollution substitution. *Journal of Environmental Economics and Management*, 102359.

Shortle, J. S., & Horan, R. D. (2001). The economics of nonpoint pollution control. *Journal of economic surveys*, *15*(3), 255–289.

von der Goltz, J., & Barnwal, P. (2019). Mines: The local wealth and health effects of mineral mining in developing countries. *Journal of Development Economics*, *139*, 1–16.

Yang, D. (2008). Can enforcement backfire? Crime displacement in the context of customs reform in the Philippines. *The Review of Economics and Statistics*, *90*(1), 1–14.

Appendix A: Mathematical Appendix

Substituting x_L, l_L^s, and $M_L(\gamma_l)$ in the utility function of low-wealth households:

$$
\begin{aligned}
u(x_L, l_L^s, H_L) = &\log\left[\frac{1}{c_x}\left[\frac{w(2-\beta)+\beta I_L + c_M(\mathcal{H}-D_L(\gamma_l))}{c_M+2}\right]\right] \\
&+\log\left[1-\left[\frac{w(c_M+\beta)-\beta I_L + c_M(D_L(\gamma_l)-\mathcal{H})}{w(c_M+2)}\right]\right] \\
&+\log\left[\mathcal{H}_L+\beta\left[\frac{\beta(I_L+w)+2(w+D_L(\gamma_l)-\mathcal{H})}{\beta(c_M+2)}\right]-D(\gamma_l)\right]
\end{aligned}
$$

Taking the derivative of $u(\cdot)$ with respect to γ_l:

$$
\begin{aligned}
\frac{\partial u}{\partial \gamma_l} = &\underbrace{-\frac{c_M D'(\gamma_l)}{w(2-\beta)+\beta I_L + c_M(\mathcal{H}-D(\gamma_l)} - \frac{c_M D'(\gamma_l)}{w(c_M+2)-w(c_M+\beta)+\beta I_L + c_M(\mathcal{H}-D(\gamma_l))}}_{(*)} \\
&\underbrace{-\frac{c_M D'(\gamma_l)}{\mathcal{H}(c_M+2)+\beta(I_L-w)+2(w+D(\gamma_l)-\mathcal{H})-D(\gamma_l)(c_M+2)}}_{(**)}
\end{aligned}
$$

Incomplete regulation causing an increase in γ_l will be regressive if $\frac{\partial u}{\partial \gamma_l} < 0$

First, let's obtain conditions under which (*) is negative:

(*) can be rewritten as:

$$
(*) = \frac{-[4c_M w - 2\beta c_M w + 2c_M \beta I_L + 2c_M^2(\mathcal{H}-D(\gamma_l))]}{w^2(4-3\beta^2)+4\beta w I_L + \beta^2 I_L^2 + c_M(\mathcal{H}-D(\gamma_l))^2 + (\mathcal{H}-D(\gamma_l))(4wc_M - 2c_M w\beta + 2\beta I_L}
$$

Recall that $0 < \beta \leq 1$, which means that the denominator of (*) will only be negative (implying the possibility of $\frac{\partial u}{\partial \gamma_l} > 0$) if the denominator of (*) is negative.

Working of the denominator of (**):

$$
(**) = w^2(4-3\beta^2)+4\beta w I_L+\beta^2 I_L^2+c_M(\mathcal{H}-D(\gamma_l))^2+(-D(\gamma_l)(4wc_M+2c_M w\beta-2c_M w\beta+2\beta I_L)
$$

(**) would only be negative if:

$$4wc_M + 2c_M w\beta - 2c_M w\beta + 2\beta I_L < 0 \implies \frac{c_M(2-\beta)}{\beta} > \frac{I_L}{w}$$

Therefore, $\frac{\partial u}{\partial \gamma_l} < 0$ unless $\frac{c_M(2-\beta)}{\beta} < \frac{I_L}{w}$.

Substituting x_H, l_H, and $M_H(\gamma_k)$ in the utility function of high-wealth households:

$$u(x_H, l_H, H_H) = \log\left[\frac{1}{c_x}(I_R + \phi Y)\right] + \log\left[\mathcal{H} + \beta\left[\frac{1}{2c_M}(I_R + \phi Y) - \frac{1}{\beta}(\mathcal{H} - D(\gamma_k))\right]\right]$$

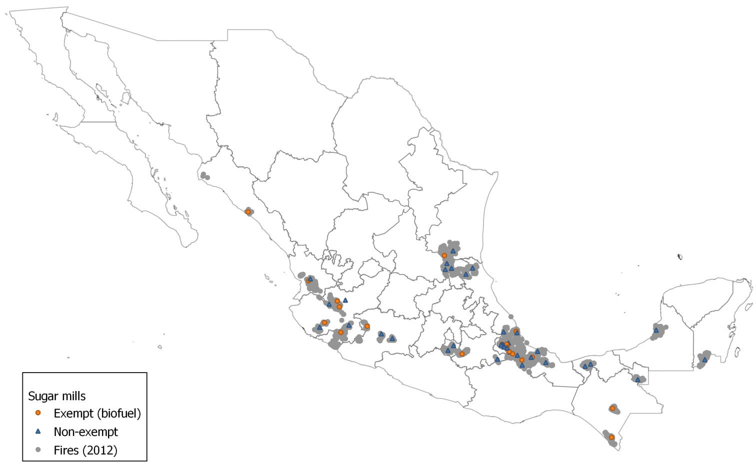
Obtaining the derivative with respect to γ_k:

$$\frac{\partial u}{\partial \gamma_k} = \frac{2c_M \frac{\partial D(\gamma_k)}{\partial \gamma_k}}{\beta(I_R + \phi Y) + 2D(\gamma_k)c_M}$$

The numerator of the previous expression is positive while the denominator will be positive unless $\beta(I_R + \phi Y) < 2D(\gamma_k)$.

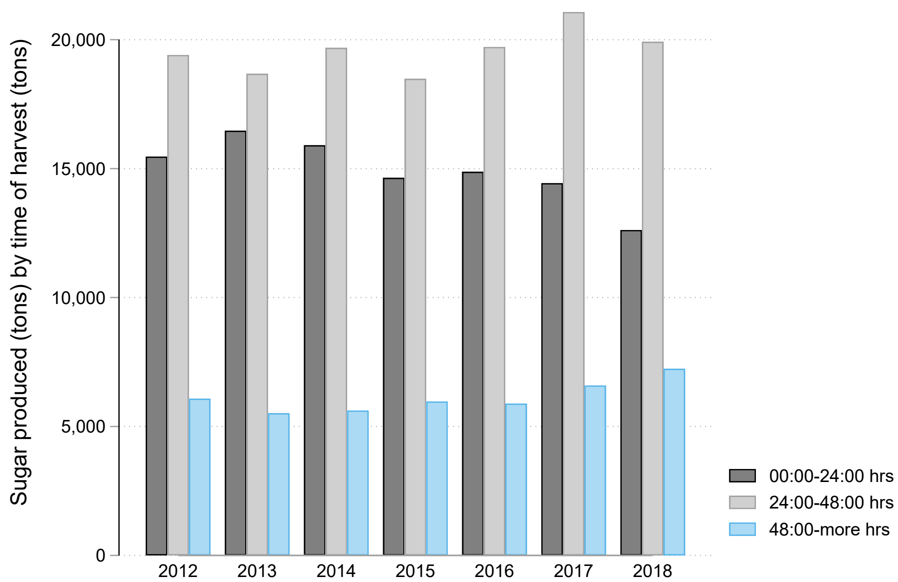
Appendix B: Additional Figures

Figure A1: Mills location



Notes: location of mills classified by exempt (biofuel use) and non-exempt or regulated by NOM-085-SEMARNAT-2011. Shadow areas are the centroid of the sugarcane fires in 2012.

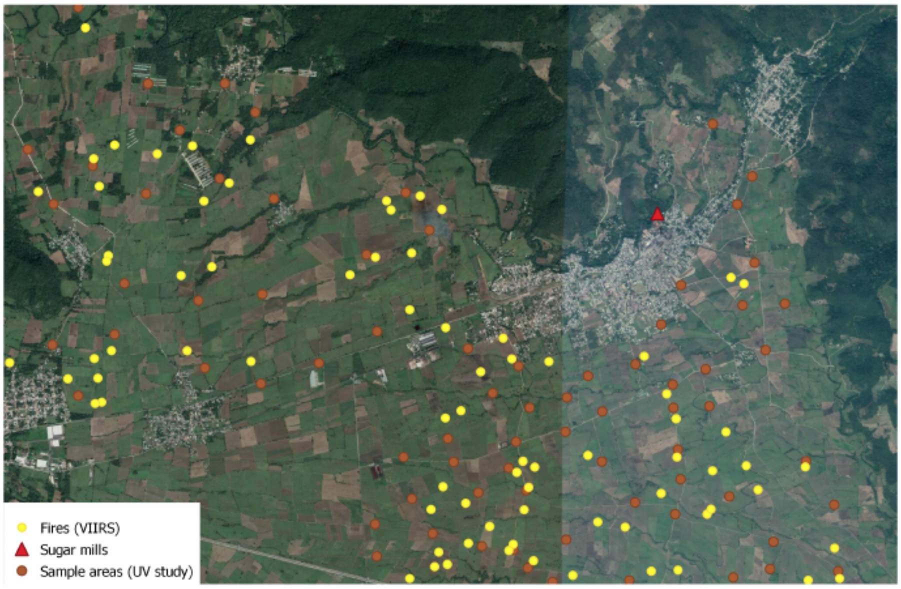
Figure A2: Time to process harvested sugarcane



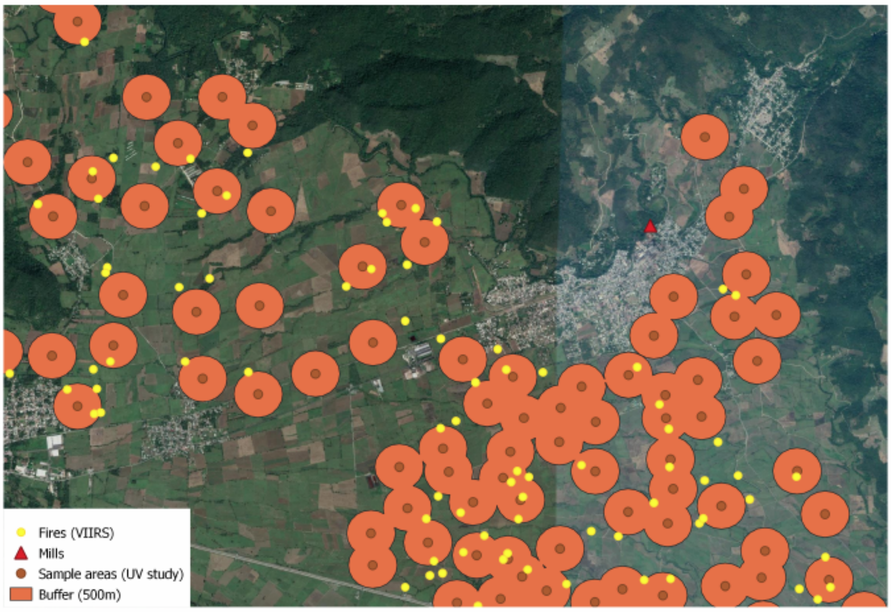
Notes: The figure shows the sugarcane processed (in tons) and its approximate time for processing in the mill after it was harvested by year.

Figure A3: Distribution areas

Panel a)

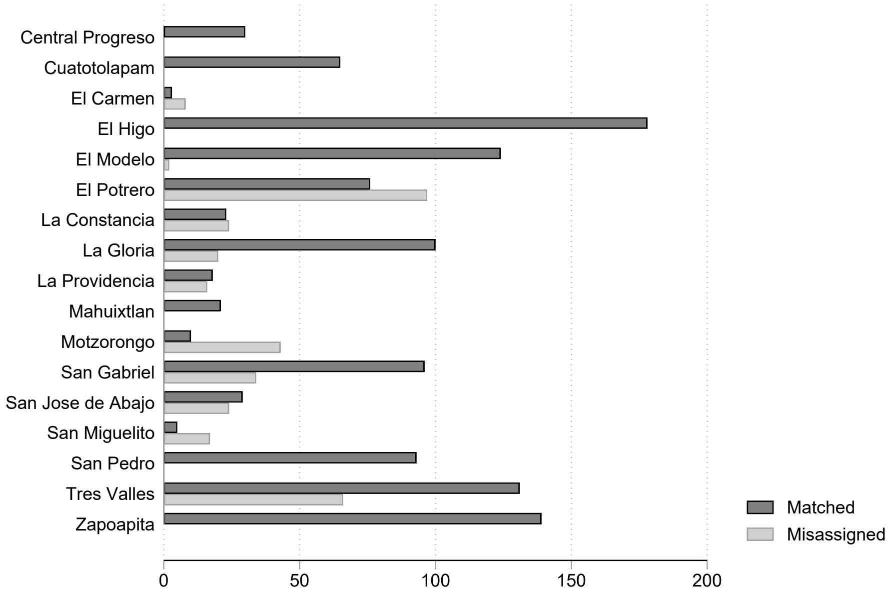


Panel b)



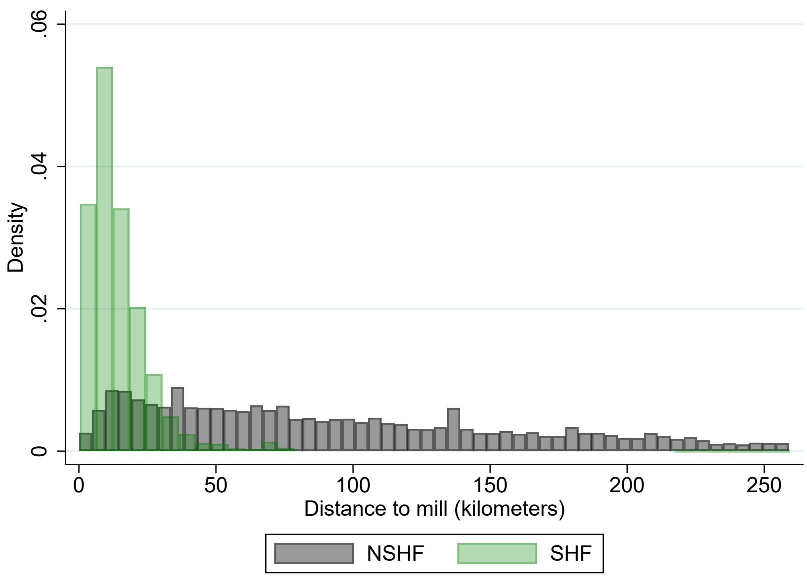
Notes: Panel a) shows the geographic extent of the sample areas (obtained by the Universidad Veracruzana study), sugarcane harvest fires (VIIRS), and a sugar mill. Panel b) shows the geographic extent of a 500m buffer surrounding the sample areas obtained by the Universidad Veracruzana study, along with the sugarcane harvest fires, and the sugar mill. The geographic area is in the city of Cordoba, Veracruz. The sugar mill is "El Potrero".

Figure A4: Distribution areas



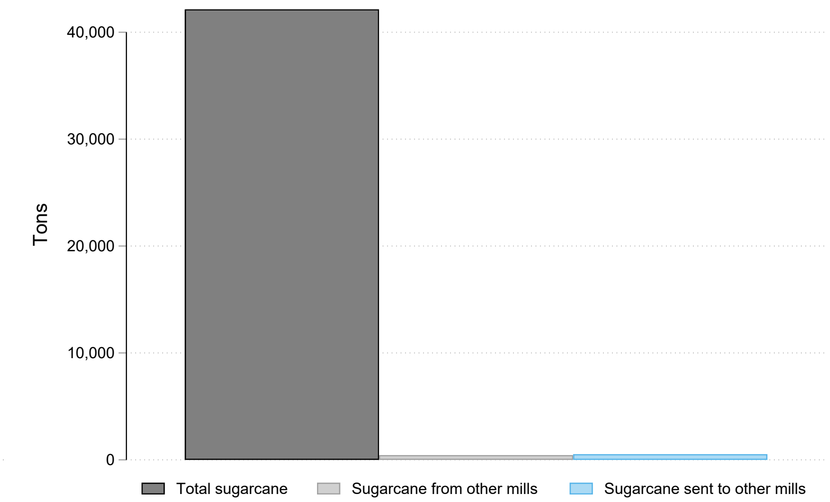
Notes: The figure shows the total number of fires for the sample area in the state of Veracruz by mill. Each bar shows the total of matched or misassigned fires comparing the distance algorithm classification and the actual sample areas for 2009.

Figure A5: Calculated distance between sugarcane fires and other fires and mills



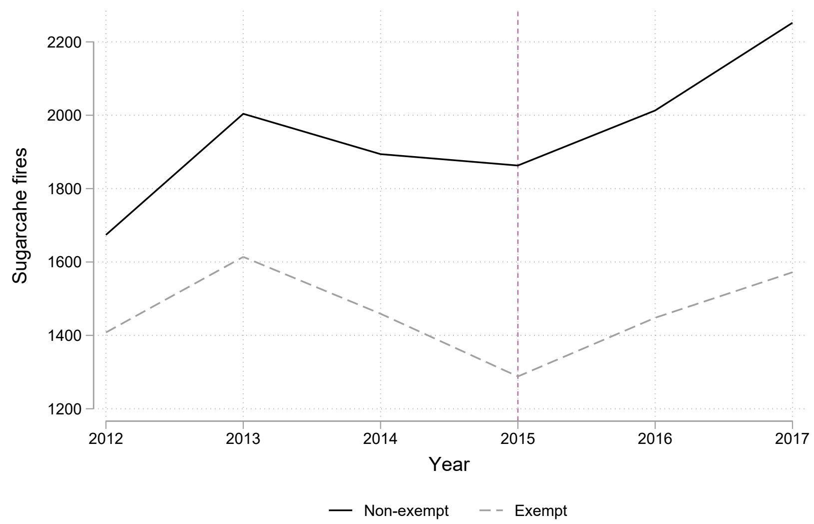
Notes: The figure shows the distribution of the distances between the fires and assigned mills. NSHF stands for "Non-sugarcane harvest fire" which are agricultural fires in cropland other than sugarcane. NSHF were classified using data from INEGI: *Cartas de Uso de Suelo y Vegetación, Serie V*. SHF stands for sugarcane harvest fire. SHF were classified by CONADESUCA and provided to the researcher.

Figure A6: Sugarcane processed by ownership



Notes: Total sugarcane processed (in tons) by type of land ownership. Total sugarcane is the sum of sugarcane from own/associated fields, sugarcane from other mills and sugarcane sent to other mills. Sugarcane from other mills is the total sugarcane sent from the fields owned/associated to other mills. Sugarcane sent to other mills is the average sugarcane that is sent to other mills due to capacity constraint associated with the own mills operations.

Figure A7: Fires associated with exempt and non-exempt facilities

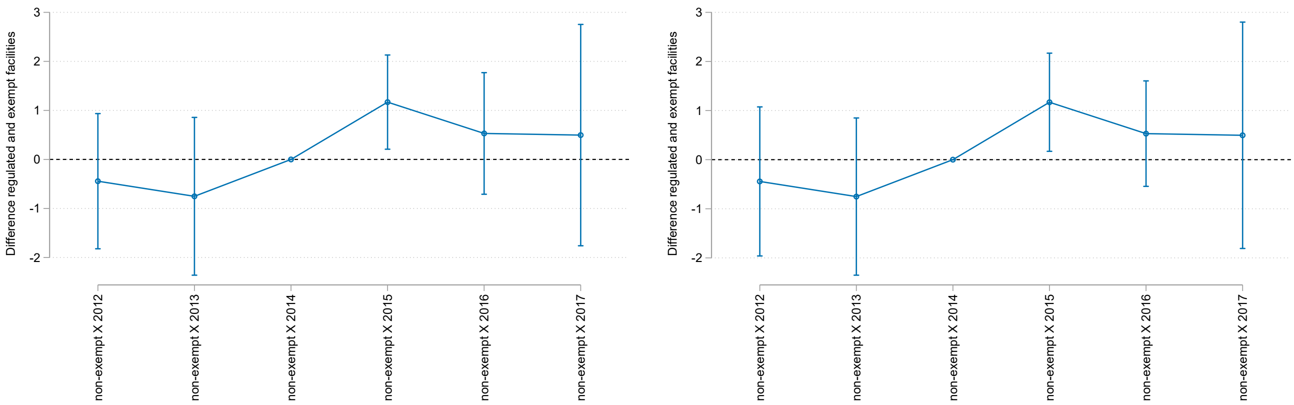


Notes: The figure show the mean sugarcane-harvest fires by type of facility. Non-exempt facilities are defined as facilities using oil as main fuel. Exempt facilities are defined as facilities using biofuels as main fuel. Vertical axis is the mean fires in sugarcane plots.

Figure A8: Monthly fires-parallel trends controlling for sugar and oil prices

(a) Controls for international price of sugar)

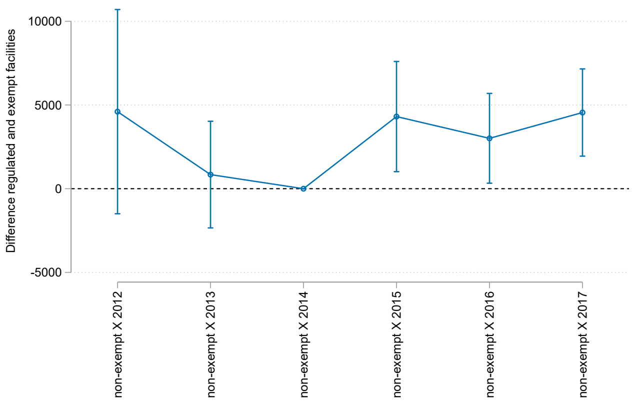
(b) Controls for the price of Mexican crude oil



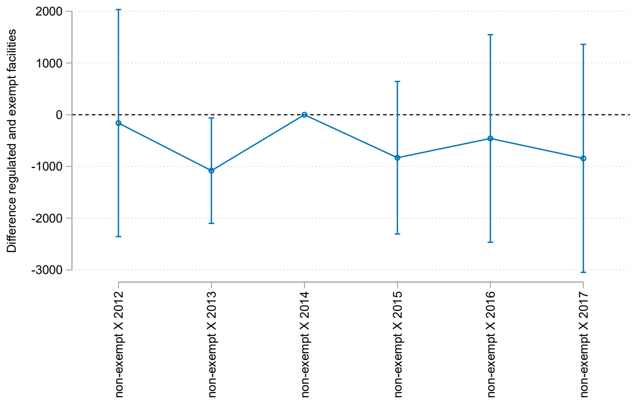
Notes: This figure shows the differences in differences-year specific coefficients for the total number of daily fires following equation 6 including controls of international sugar prices (Panel a)) and Mexican crude oil prices (Panel b)). The regulation started in 2015. 95% confidence intervals calculated using two-way fixed effects at the municipality and year level.

Figure A9: Inputs-parallel trends

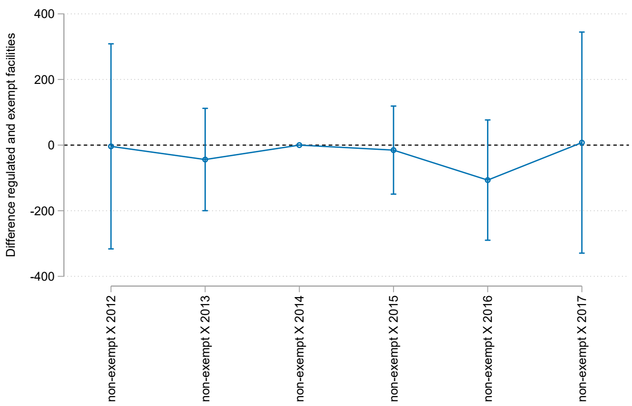
(a) Manual cut (tons)



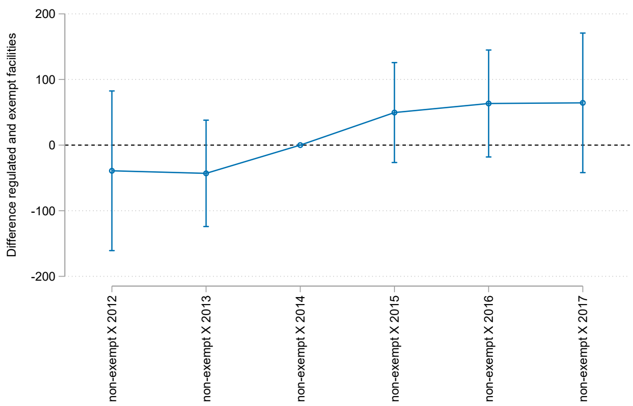
(b) Mechanical cut (tons)



(c) Hectares harvested



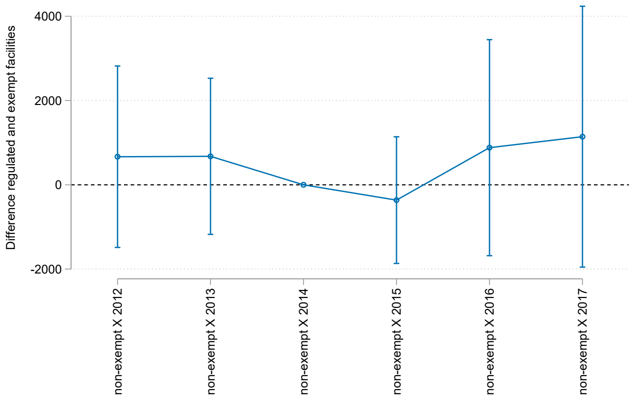
(d) Total workers



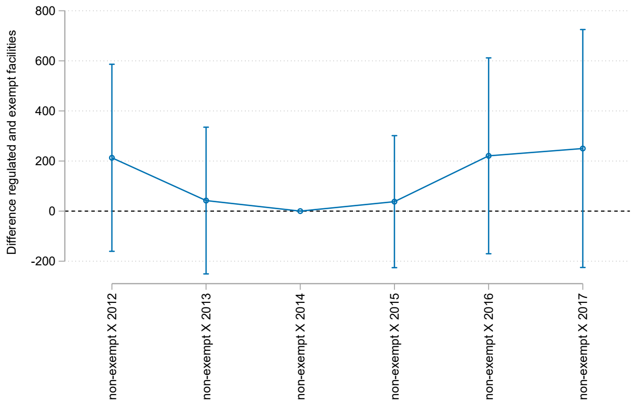
Notes: Panel a) shows the differences in differences-year specific coefficients for the total sugarcane in tons using manual cut following equation 7. Panel b) shows the differences in differences-year specific coefficients for the total sugarcane in tons using mechanical cut following equation 7. Panel c) shows the differences in differences-year specific coefficients for the harvested hectares following equation 7. Panel d) shows the differences in differences-year specific coefficients for the total number of manual workers following equation 7. Confidence intervals calculated using two-way fixed effects at the municipality and year level.

Figure A10: Outputs-parallel trends

(a) Processed sugarcane



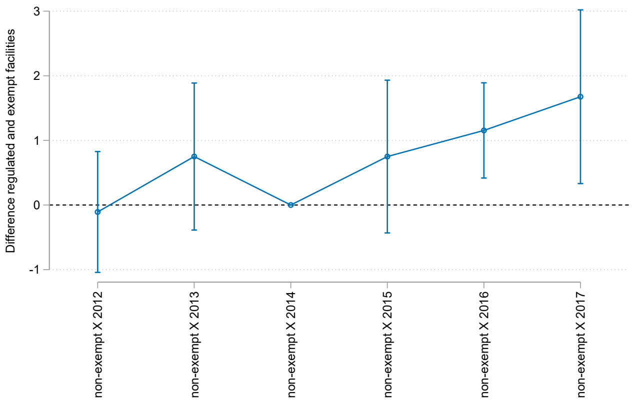
(b) Total sugar produced



Notes: Panel a) shows the differences in differences-year specific coefficients for the total sugarcane processed at the mill (mills) following equation 7. Panel b) shows the differences in differences-year specific coefficients for the total produced sugar at the mill (tons) using mechanical cut following equation 7. Confidence intervals calculated using two-way fixed effects at the municipality and year level.

Figure A11: Pollution in fields-parallel trends

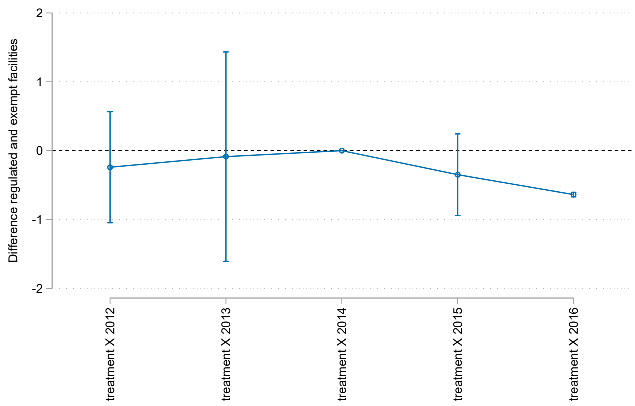
(a) $PM_{2.5}$ pollution ($\mu g/m^3$)



Notes: Panel a) shows the differences in differences-year specific coefficients for pollution from $PM_{2.5}$ in the fields following equation 8. Panel b) shows the differences in differences-year specific coefficients for pollution from SO_2 in the fields following equation 8. Confidence intervals calculated using two-way fixed effects at the municipality and year level. Confidence intervals calculated using two-way fixed effects at the municipality and year level.

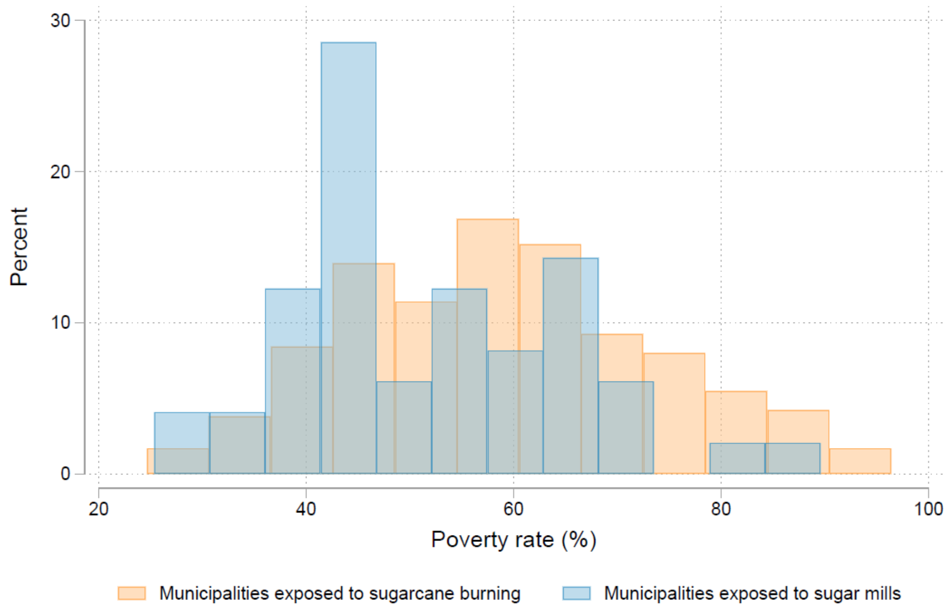
Figure A12: Pollution in mills-parallel trends

(a) $PM_{2.5}$ pollution ($\mu g/m^3$)



Notes: Panel a) shows the differences in differences-year specific coefficients for pollution from $PM_{2.5}$ in the mills following equation 8 using Hammer et al. (2020). Confidence intervals calculated using two-way fixed effects at the municipality and year level. Confidence intervals calculated using two-way fixed effects at the municipality and year level.

Figure A13: Poverty and location of fires and mills



Notes: The figure shows the percent of communities exposed to sugarcane burning and sugar production and their corresponding poverty rate. Poverty rate was obtained from INEGI using the 2010 Mexican Census data.

Appendix C: Additional Tables

Table A1: Descriptive statistics by type of facility before the policy

	(1)	(2)
	Exempt facilities	Non-exempt facilities
	Panel a: Inputs data	
Manual cut (tons)	27769.69	40982.75
	(15268.95)	(24692.98)
Mechanical cut (tons)	5796.26	8880.16
	(6671.87)	(7900.13)
Total field workers (cutting)	1325.13	1811.07
	(650.75)	(1116.84)
Total harvested sugarcane (hectares)	327.90	445.49
	(357.74)	(618.76)
Total sugarcane (tons)	3041.26	3015.94
	(4286.52)	(3887.12)
Sugar production efficiency	98.16	101.40
	(44.40)	(40.16)
Sugarcane processing efficency	111.59	111.83
	(29.02)	(25.07)
Sugar extraction efficiency	20.34	13.12
	(281.80)	(2.29)
	Panel b: Outputs data	
Raw processed sugarcane (t)	36006.08	51574.08
	(15837.38)	(23785.49)
Total sugar produced (t)	3945.55	5901.07
	(1898.63)	(2857.84)
Total sugar produced per day (t)	555.64	828.56
	(269.05)	(399.47)
Observations	3,006	1,517
	Panel c: Fires and temperature data	
Total SHFs	0.25	0.37
	(0.75)	(0.95)
Temperature (C)	21.85	21.60
	(3.78)	(3.17)
Observations	47,690	23,845

Notes: Panel a) shows descriptive statistics for inputs and panel b) shows descriptive statistics for outputs at the sugar mill level at the weekly level. Panel c) shows descriptive statistics of sugarcane harvest fires and average temperature for daily observations at the sugar mill level. Column (1) shows the descriptive statisitcs for exempt facilities and Column (2) shows the descriptive statistics for non-exempt facilities. Standard deviation in parentheses.

Table A2: Difference-in-differences estimates for monthly fires

	(1) Total SHFs	(2) SHF=1	(3) Log(SHF)
After 2015 × non-exempt	1.12952***	0.00133	0.11222***
	(0.38957)	(0.01743)	(0.03168)
Pre 2015 mean	8.558	0.825	1.870
Obs.	2,394	2,394	1,975
R-squared	0.524	0.450	0.573
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Linear trend	No	No	No
Weather Controls	Yes	Yes	Yes
Cluster level	Mun and year	Mun and year	Mun and year
Poisson	No	No	No

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of fires using equation 6. Column (2) shows the same specification but the dependent variable is an indicator variable on whether there is a fire or not in that field. Column (3) estimates the same specification in equation 6 with the log number of fires at the month level. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A3: Difference-in-differences estimates for monthly fires controlling for sugar and oil prices

	(1) Total SHFs	(2) Total SHFs
After 2015 × non-exempt	1.13018**	1.12949**
	(0.38912)	(0.39060)
Pre 2015 mean	8.558	8.558
Obs.	2,394	2,394
R-squared	0.525	0.524
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mun and year	Mun and year
Additional controls	Sugar price	Oil price

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of fires using equation 6 controlling for the international sugar prices. Column (2) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of fires using equation 6 controlling for the Mexican mix crude oil prices. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A4: Effect on sugarcane fires using bootstrap standard errors

	(1) Total SHFs
After 2015 × non-exempt	0.04120**
	[0.00982,0.07258]
Pre 2015 mean	0.286
Obs.	71,535
R-squared	0.091
Year FE	Yes
Month FE	Yes
Mill FE	Yes
Weather controls	Yes
Cluster level	Bootstrap
Poisson	Yes

Notes: The table shows the results from equation (6) using bootstrap standard errors. Column (1) shows the resulting specification with dependent variable being total number of fires. Confidence intervals reported in brackets.

Table A5: Effect on non-sugarcane fires

	(1) Total NSHFs	(2) NSHF=1
After 2015 × non-exempt	0.24180	-0.01129
	(0.25467)	(0.00806)
Pre 2015 mean	3.060	0.439
Obs.	71,535	71,535
R-squared	0.321	0.387
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mun and year	Mun and year

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of agricultural fires (non-sugarcane) using equation 6. Non-sugarcane fires were classified using land cover data from INEGI. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A6: Placebo test: sugarcane fires outside the harvest season

	(1) Total SHFs	(2) SHF=1
After 2015 × non-exempt	0.00218	0.00046
	(0.00590)	(0.00389)
Pre 2015 mean	0.034	0.024
Obs.	52,326	52,326
R-squared	0.054	0.059
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mun and year	Mun and year

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of fires using equation 6 for the months July-October, outside the sugarcane harvest window. Column (2) shows the same specification but the dependent variable is an indicator variable on whether there is a fire or not in that field. Column (3) estimates the same specification in equation 6 with the log number of fires at the month level. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A7: Effect on sugarcane fires - distribution areas

Panel a)

	(1) Total SHFs	(2) SHF=1	(3) Total SHFs
After 2015 × non-exempt	0.02284*	0.01423**	0.32656***
	(0.01111)	(0.00527)	(0.11845)
Pre 2015 mean	0.067	0.051	0.067
Obs.	25,000	25,000	25,000
R-squared	0.027	0.031	
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Cluster level	Mun and year	Mun and year	Robust
Poisson	No	No	Yes
Buffer	300	300	300

Panel b)

	(1) Total SHFs	(2) SHF=1	(3) Total SHFs
After 2015 × non-exempt	0.05044*	0.01985**	0.24923**
	(0.02502)	(0.00710)	(0.12634)
Pre 2015 mean	0.170	0.086	0.170
Obs.	25,080	25,080	25,080
R-squared	0.039	0.051	
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Cluster level	Mun and year	Mun and year	Robust
Poisson	No	No	Yes
Buffer	500	500	500

Notes: Panel a) and b) show the results following the equation 6 considering the sugarcane distribution areas for the subset of sugar mills in Veracruz under the study " Digitalización del Campo Cañero en México para Alcanzar la Agricultura de Precisión de la Caña de Azúcar". Panel a) shows the results using a buffer of 300m surrounding the sampling points and Panel b) shows the results using a buffer of 500m surrounding the sampling points. Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of sugarcane fires. Column (2) shows the same specification but the dependent variable is an indicator variable on whether there is a fire or not in that field. Column (3) estimates the specification using a Poisson model using robust standard errors. Standard errors using two way clusters (munincipality and year) in parenthesis for Columns (1) and (2).

Table A8: Effect on sugarcane fires - compliant mills

	(1) Total SHFs	(2) SHF=1
After 2015 × non-exempt	0.09283**	0.02223*
	(0.03238)	(0.01011)
Pre 2015 mean	0.314	0.177
Obs.	13,805	13,805
R-squared	0.120	0.135
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Cluster level	Mun and year	Mun and year
Compliant	Yes	Yes

Notes: The table shows the main results for daily fires with a restricted sample for facilities with known compliance by CONADESUCA. Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the number of sugarcane harvest fires using equation 6. Column (2) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the probability of sugarcane fire. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A9: Effect on sugarcane fires - compliant mills

	(1) Sample restriction 1	(2) Sample restriction 2
After 2015 × non-exempt	0.03935**	0.04077**
	(0.01505)	(0.01540)
Pre 2015 mean	0.285	0.289
Obs.	70,280	70,280
R-squared	0.092	0.091
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mun and year	Mun and year
Poisson	No	No

Notes: The table shows the main results for daily fires with two restricted sample of facilities. Column (1) shows the results with the sample restriction 1 that estimates the regressions dropping the mill "San Francisco Ameca" which is the only mill that acquired a biofuel powered boiler after the treatment started. Column (2) shows the results with the sample restriction 2 that estimates the regressions dropping the mill "Ingenio El Potrero", which is the only mill that has the highest rate of mismatched fields according to Figure A4. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A10: Effects on wages

	(1)	(2)	(3)	(4)	(5)
			Wage		
After 2015 × non-exempt	-1.770	-5.172	-1.236	1.722	-0.308
	(6.57693)	(8.05965)	(8.66429)	(12.50082)	(5.40046)
Mean	103.484	109.047	113.478	112.093	94.440
Obs.	9,769	1,978	2,005	2,001	1,969
R-squared	0.530	0.719	0.780	0.716	0.644
Year FE	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes	Yes	Yes
Cluster level	Mun-year	Mun-year	Mun-year	Mun-year	Mun-year
Age	All	15-30	30-45	45-60	60-more

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the daily wages following 7. Column (2)-(7) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the wages for the workers different age categories. Data obtained from the social security information at the municipality level. The sample of mills was restricted to the municipalities that have either all mills with the same treatment status (all exempt or all non-exempt), therefore deleting 4 mills. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A11: Effects on outputs

	(1)	(2)
	Raw processed sugarcane (tons)	Total sugar produced (tons)
After 2015 × non-exempt	83.171	78.483
	(904.27822)	(138.69765)
Mean	41,865.426	4,674.832
Obs.	8,568	8,568
R-squared	0.803	0.792
Year FE	Yes	Yes
Week FE	No	No
Mill FE	Yes	Yes
Cluster level	Mun and year	Mun and year

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the amount of sugarcane processed at the mill (tons) following specification 7. Column (2) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the amount of sugar produced (tons) following specification 7. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A12: Effects on efficiency

	(1)	(2)	(3)
	Sugarcane processing efficency	Sugar production efficiency	Sugar extraction efficiency
After 2015 × non-exempt	-0.857	-4.678	-7.528
	(4.79311)	(6.09324)	(6.40081)
Mean	113.734	101.143	15.276
Obs.	5,707	5,707	5,707
R-squared	0.425	0.615	0.020
Year FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Cluster level	Mun and year	Mun and year	Mun and year

Notes: Column (1) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the total kilograms of sugar obtained by ton of harvested sugarcane (measured by the KARBE indicator provided by CONADESUCA) at the mills following specification 7. Column (2) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the total kilograms of sugar obtained by ton of processed sugarcane (measured by the KABE indicator provided by CONADESUCA) following specification 7. Column (3) shows the difference-in-differences estimator of the impact of being regulated by the emission limits after the policy started on the sugar extraction efficiency. Standard errors using two way clusters (munincipality and year) in parenthesis.

Table A13: Effect on sugarcane fires by poverty level

	(1) Total SHFs	(2) Total SHFs	(3) Total SHFs
After2015=1=1 × Treatment=1=1	0.03278***	0.02212	0.04010***
	(0.00653)	(0.02501)	(0.00349)
Mean	0.289	0.266	0.312
Obs.	64,005	31,375	32,630
R-squared	0.087	0.074	0.096
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Linear trend	No	No	No
Weather controls	No	No	No
Cluster level	Mun and year	Mun and year	Mun and year
Poverty level (wrt median)	All	Lower	Higher

Notes: The table shows the results from equation (6) with heterogeneity by poverty level in fields areas. Column (1) shows the resulting specification with dependent variable being total number of fires. Column (2) shows the results for fields located near localities with poverty level lower than the national median. Column (3) shows the results for fields located near localities with poverty level higher than the national median. Twoway-clustered standard errors