The Distributional Consequences of Incomplete Regulation

Danae Hernández-Cortés†

September 2022

Abstract

Incomplete environmental regulation can shift production from regulated to unregulated sectors, affecting the spatial distribution of pollution and who bears its burden. I study this phenomenon in the context of sugarcane processing in Mexico. Firms responded to requirements to install air pollution controls in mills by increasing agricultural fires in sugarcane fields by 15%. As a result, $PM_{2.5}$ concentrations rose by 7% with higher impacts in socioeconomically vulnerable communities. These findings highlight an often undiscussed implication of incomplete pollution regulation: its distributional consequences.

†Arizona State University, Danae.Hernandez-Cortes@asu.edu. I appreciate detailed feedback and support from 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, Nicholas Flores, Don Fullerton, Matthew Gibson, Robert Heilmayr, Juliana Helo, Sarah Jacobson, Bryan Leonard, Antony Millner, Andrew Plantinga, Paulina Oliva, Chris Severen, Glenn Sheriff, Eric Verhoogen, and members of the UCSB Environmental and EnvDev Reading Groups.

Introduction

Many environmental policies regulate pollution-generating activities instead of regulating pollution directly. As a result, production can be reallocated across locations and within supply chains, escaping the reach of regulation. In such circumstances, the regulation is "incomplete" and firms may substitute production from a regulated to an unregulated activity. Incomplete regulation has well established efficiency consequences (Fowlie, 2009; Gibson, 2018).[1] However, much less attention has been paid to whether incomplete pollution policies have distributional consequences. For example, individuals living near regulated activities might experience relative decreases in pollution as a result of regulation while individuals near unregulated activities experience higher pollution levels.

This paper analyzes 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 economic activity in southern and central Mexico. While economically important, sugarcane harvest is a heavily polluting activity. The fact that there are two primary harvest methods, mechanical and manual cut, provides the potential for significant pollution reallocation under incomplete regulation.

These harvest alternatives and the regulatory history of sugarcane production in Mexico provide an interesting setting for studying the distributional consequences of incomplete regulation. As already mentioned, sugar mills have two technological alternatives for harvesting sugarcane: mechanical or manual cut. When sugarcane is cut manually, it is often burned since fires clean excess vegetation on the sugarcane plant. These fires have been found to affect in-utero health outcomes for nearby populations (Rangel & Vogl, 2019). If sugarcane is not cleaned in the field using fires and is instead mechanically cut, it needs to go through an additional cleaning process that uses industrial boilers. Starting in 2015, the Mexican government implemented a policy aimed at decreasing sulfur dioxide emissions from industrial boilers, requiring facilities in all sectors of the

[1] Incomplete regulation and leakage have been studied in several settings such as international trade, where countries export pollution to countries with laxer pollution regulation (Levinson & Taylor, 2008; Copeland & Taylor, 2004), climate policies (J. B. Bushnell & Mansur, 2011; J. Bushnell & Chen, 2012; Perino, 2015; Baylis et al., 2013; Fowlie & Reguant, 2020), and other environmental regulations (Becker & Henderson, 2000; Chan & Morrow, 2019). Most leakage studies focus on analyzing the reallocation of the same pollutant or the same economic activity in different areas. These studies find that under incomplete regulation, firms substitute towards unregulated areas, offsetting or even increasing total emissions compared to regulated areas.

economy using oil as fuel to reduce emissions each year by either substituting to less polluting boilers or acquiring abatement technologies. 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 manual cutting and associated harvest fires relative to unregulated mills. Furthermore, I show that mills shifted the type of inputs used (i.e. manual labor), altering the spatial distribution of pollution and the health outcomes of the populations exposed to pollution.

I use a difference-in-differences research design to compare input use and pollution from regulated and unregulated facilities before and after the policy was implemented. I find that fields linked to regulated mills increased the number of sugarcane fires by 15% following the regulation. As a result, ambient concentrations of $PM_{2.5}$ over these fields increased by 7%. I corroborate these substitution responses to the regulation using detailed data on various sugarcane production inputs and outputs. Consistent with an increase in fires, I find that fields linked to regulated mills increased manual cut workers by 6% 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 and provide suggestive evidence of a decrease in processing efficiency after the regulation.

I further examine whether this change in pollution disproportionately affected vulnerable rural areas. Similar to other low and middle income countries, the agricultural fields in Mexico are located near rural areas that have higher levels of poverty and lower access to health services than urban areas. I analyze whether pollution concentrations increased in areas exposed to the increase in fires, and whether these areas have higher levels of economic disadvantage. I find that the most vulnerable households experienced the largest increases in pollution driven by the policy compared to less vulnerable populations. These results highlight an important finding that has not been empirically documented in the literature: incomplete regulation can contribute to environmental inequality by altering the spatial distribution and composition of pollution.

Finally, I examine whether the increase in pollution caused by incomplete regulation is associated with worse health outcomes in affected areas. I use individual birth records for the period 2012-2017 to estimate the impacts of pollution exposure on birth outcomes such as birth weight, gestational length, very low birth weight incidence ($< 1{,}500$ g), and very preterm birth incidence (< 32 weeks). I link the location of fires to the mother's community of residence and estimate the impact of pollution on birth outcomes. I find that an additional μg/m^3 increase in $PM_{2.5}$ pollution decreases birth weight by 1.7 grams, increases very low birth weight incidence by 4%, and increases very preterm birth incidence by 5% for all newborns. The negative effects of this increase in pollution is higher among socioeconomically disadvantaged households. I find a decrease of 2.48 grams, an 11% increase in very low birth weight incidence, and a 12% increase in very preterm birth incidence for newborns whose mothers are affiliated to *Seguro Popular*, the health insurance that covers informal workers and rural communities, which are among the most vulnerable populations in Mexico. 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 (2019), 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 inputs and this substitution results in higher pollution concentrations for nearby communities. Using my estimates, I calculate that an input tax of $20 USD per ton of sugarcane harvested using manual cut (average sugarcane price is $650 USD per ton) would internalize the impacts of pollution from sugarcane burning on affected populations.

This paper has two main contributions. First, I contribute to existing studies that find input substitution responses as a result of environmental regulation (Gibson, 2018; Rijal & Khanna, 2020), and the detailed firm data allow me to quantify production complementarities and their pollution consequences. 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; Tanaka et al., 2021). By focusing on one sector and using detailed production data, I am able to unravel how firms alter production processes to adjust to regulation. 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 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 examining the distributional impacts of environmental policies. Agricultural fields are mainly located in rural areas that are on average poorer and face higher socioeconomic vulnerability than their urban counterparts where mills are located. By increasing the number of fires and pollution in these areas, this incomplete command and control regulation increased 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; Banzhaf et al., 2019; Chakraborti & Shimshack, 2020). This paper finds another source of environmental injustice: incomplete regulation can cause pollution leakage to vulnerable populations even when a command and control policy targets the reduction of pollution from every regulated facility. This contributes to existing discussions over the environmental justice consequences of environmental policies and the role of the design of these policies exacerbating or decreasing environmental inequities (Fullerton & Muehlegger, 2019; Currie et al., 2020; Holland et al., 2019; Hernandez-Cortes & Meng, 2020). 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 communities.

The rest of the paper proceeds as follows. Section 1 describes the sugarcane sector and boiler regulation. Section 2 describes the data. Section 3 presents the empirical specification. Section 4 explains the effects of regulation-induced pollution redistribution and its distributional consequences as well as the health effects of incomplete regulation. Section 5 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 alternatives for harvesting sugarcane: either mechanical or manual cut. When sugarcane is manually cut, it must also be burned in the field. 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] Mills generally own the fields where they source the sugarcane which means that mill management has a decision power over the type of harvesting alternative used.[3] It is important to note that there is no quality difference between sugarcane manually or mechanically cut. 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 that uses equipment fueled by boilers. These boilers can either use diesel, fuel oil, biofuels, or natural gas. Once clean, machines grind and extract the caloric content of sugarcane to then crystalize and refine the sugar in the mill.

The sugar producing industry is an important economic sector in 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 has health implications: sugarcane fires increase particulate matter (PM), nitrogen oxides (NO_x), and carbon monoxide (CO) concentrations. 96% of these particles are respirable (França et al., 2012) and have been associated with negative health outcomes in nearby communities. For instance, Rangel and Vogl (2019) found that in utero exposure to pollution from sugarcane fires reduces

[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 of 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.

[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.

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 emissions related to 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 sulfur dioxide, 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, unregulated facilities could respond by either complying with the regulation or by shifting technologies in the field to decrease the emissions coming from the regulated technology.

[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 (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.

[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.

The regulation is enforced by the Mexican Environmental Ministry's regulator entity, PROFEPA, and 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. Although biofuels are exempt from the regulation, 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] 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 authorities if the fire grows uncontrollably. However, sugarcane harvest fires are not regulated.

Based on this regulation background, Appendix A develops a conceptual framework to derive possible producers' responses to regulation and impacted households. A producer decides the optimal amount of each input, and generates pollution associated with the use of either of the two inputs. A social planner implements an emissions tax to only one input (i.e. incomplete regulation). This conceptual framework has two 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 defensive expenditures 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.

These theoretical predictions have implications for the setting studied in this paper. Under the new regulation of sugar mills, we might 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 rest of the paper leverages variation induced by the regulation to compare regulated and unregulated

[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).

facilities before and after the regulation was introduced and the distributional and health implications of this incomplete regulation.[9]

2 Data

I use 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 create 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.

2.1 Fires data

I obtain 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.[10] I restrict the fires to the November to May period to cover the sugarcane harvest season.[11]

2.2 Sugarcane coverage data

In order to identify the extent of sugarcane fields in Mexico, I use data from Mexico's National Committee for Sugarcane Suistainable Development (CONADESUCA). The data include confidential information of sugarcane plots in Mexico. CONADESUCA uses Landsat 8 images from 2014-2015 to map the sugarcane plots in Mexico. To do so, they

[9] One concern using this specification is that unregulated facilities could change regulation status after the policy started in order to be exempt from 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 on the three years (short run) of the policy. Moreover, I obtained data on all of the industrial investments done by mills 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 investments done by mills occurred during 1980-2000.

[10] 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.

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

classify sugarcane fields using the Normalized Difference Vegetation Index (NDVI) and validate them using Landsat data and field visits.[12] Given the confidentiality of the data, CONADESUCA links the fires centroids from VIIRS with the sugarcane fields polygons for this project. Therefore, I am able to identify whether a particular fire event happens inside a sugarcane polygon. I obtain a total of 23,106 sugarcane fires for the study period 2012-2017.[13] These fires are classified as sugarcane-harvest fires.

2.3 Sugar mills characteristics

I obtain the geographic location of all sugar mills from from the National Statistical Directory of Economic Units from INEGI, based on the Economic Census 2009 performed by INEGI. Figure A1 shows the geographic coverage of the sugar mills along with their regulation classification based on fuel use pre-policy. The fuel use pre-policy comes from CONADESUCA sustainability annual reports. A mill is considered to be part of the regulated group if it did not use biofuels in their production process or if it did not use biofuels to co-generate electricity for its production activities during 2010 and 2011.[14]

2.4 Production inputs and outpus

I add the mills' location information to detailed mill-level production data from the Sinfocaña system updated by CONADESUCA.[15] This 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 using manual and mechanical cut, hours worked, among other field information.[16]

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

[13] In order to correct for measurement error between the VIIRS resolution and the sugarcane fires provided by CONADESUCA, I also create a 50 m buffer around the fires and classify 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 is done in order to account for fires that are not classified as sugarcane fires. This procedure yields a total of nearly 200 additional sugarcane fires.

[14] 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. I cross-validated the regulated vs. unregulated definition using a list of compliance at the mill level provided by PROFEPA.

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

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

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 regulation (NOM-085-SEMARNAT-2011).[17] Table A1 shows descriptive statistics for facilities using biofuels (unregulated) and oil (regulated). Unregulated 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.

I also obtain agricultural daily wages for employed workers at the municipality level which is based in payroll contributions to the Social Security Institute (*Instituto Mexicano del Seguro Social*, IMSS). The data contain total workers by sector, age, and gender at the municipality level. IMSS covers mainly formal workers which 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. Payroll data are 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 (2 out of 45 municipalities, four sugar mills total).

2.4.1 Linking fires and sugar mills

I link sugar mills to their distribution fields by calculating the distance from the sugarcane fires to all existing mills and assigning fires to the closest mill. Distance to the mill is likely a good indicator of ownership: sugarcane needs to be processed within a week of being harvested or it can lose caloric content and produce less sugar.[18] 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 incur in the transportation costs from the fields using their own trucks and lend mechanical har-

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

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

vest technologies to smallholders.[19] To the extent that there are no consistent differences between regulated and unregulated facilities in misassignment of the fires, measurement error linking facilities and mills is likely to downward bias my estimates.

I corroborate the link by obtaining information on a random sample of the supply fields for the mills in the state of Veracruz, the largest sugarcane producer in the sate, and calculate the overlap of the sugarcane harvest fires definition with these fires. The data contain sampling points of sugarcane fields in the state of Veracruz and was obtained by the Universidad Veracruzana in 2009.[20] I calculate a buffer of 300 and 500 meters surrounding the sampling sites and compared the mills association of these fires. I find 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.[21] 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 compared to the full set of fires, consistent with attenuation bias from classical measurement error.[22]

2.4.2 Other agricultural fires

If incomplete regulation increases manual cut and sugarcane fires, other agricultural fires should not increase after the policy. To perform this falsification test, I obtain the number of agricultural fires that are not associated to sugarcane areas in order to construct a fires "placebo" group. I use the rest of the non-sugarcane fires in the VIIRS data not classified

[19]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 with sugar mill workers and they mentioned that this is a small percentage.

[20]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.

[21]One limitation of this method 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.

[22]Sugar mills often 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 A5 shows the total sugarcane harvested by the own mill and the amount either sold or received from other mills. Furthermore, 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.

as belonging to a sugarcane polygon by CONADESUCA. I classify them as agricultural fires if they are 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 follow the same procedure to classify each fire in the mill "catchment" area as for the sugarcane fires. As another robustness check, I classify fires as "non-harvest sugarcane fires" if these fires occur within a sugarcane field in the months June-October, outside the usual harvest window for sugarcane.[23]

2.5 Pollution data

I obtain daily pollution data from NASA's MERRA-2 aerosol optical depth product.[24] Daily pollution data has a $0.5\,^\circ \times 0.625\,^\circ$ resolution. This reanalysis AOD product has information on particulate matter precursors and I calculate $PM_{2.5}$ 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 link the fires and mills coordinates to the pollution pixels and calculate the pollution associated with 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 use 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 $0.01° \times 0.01°$. 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 have been used in other contexts (Fowlie et al., 2019, Garg et al., 2020) and the spatial definition is desirable to analyze detailed spatial units such as mills. The downside of the data is the temporal scale since it only provides annual estimates of particulate matter which are likely a poor measurement for seasonal pollution patterns such as agricultural fires.

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

[24] Specifically, I use the diurnal, time-averaged, single level assimilation, Aerosol Diagnosis V5.12.4 (M2TUNXAER).

2.6 Birth outcomes data

Data on birth outcomes comes 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 residence information such as number of doctor visits, age, education, employment, and community of residence.[25] I am able to link the community 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.[26] I merge average daily birth outcomes at the community to the average monthly pollution concentrations due to fires in each month of the last pregnancy trimester.

2.7 Socioeconomic characteristics

In order to analyze the distributional consequences of incomplete regulation and pollution leakage, I use 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 like the percentage of people older than 15 without education, percentage of households without piped water, bathroom, electricity, and refrigerator, average number of people living in a household, among others. The index uses data at the community 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 to calculate poverty levels at the community level in order to analyze whether poorer communities experienced a higher increase in fires.

[25]Throughout this paper, community refers to a *localidad*, which denotes a small administrative unit in Mexico (smaller than a municipality).

[26]Studies have found that pollution exposure has negative birth outcomes for the last pregnancy trimester (Currie et al., 2009; Rangel & Vogl, 2019)

3 Empirical Specification

3.1 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 compare regulated facilities (oil burning mills) and unregulated facilities (biofuel burning mills). In order to analyze whether regulated facilities substitute from cleaning sugarcane in the mills using regulated boilers to cleaning in the fields using fires, I use a difference-in-differences approach:

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

Where SHF_{imdt} is the sugarcane harvest fires on day d associated with mill i in month m and year t, 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_{idmt} are weather controls, and ϵ_{idmt} are standard errors clustered at the mill level. β_1 shows the difference-in-difference estimate of the impact of being regulated by the new emission limits after 2015.

The identifying assumption of equation (1) is that in the absence of treatment, fires in both regulated and unregulated facilities would have followed the same trend. Showing parallel trends in the outcomes of interest allows me to test for differences in the groups prior to the introduction of the policy.[27] Figure A6 Panel B shows that pre-treatment, both regulated and unregulated facilities follow similar trends in the number of daily fires. This figure suggests that prior to the start of the program, fires in regulated and unregulated facilities followed a similar trend. I perform several robustness checks, including controlling for international sugar prices and Mexican crude oil prices.[28] I also perform two falsification tests. First, I replace the dependent variable with $NSHF_{idmt}$

[27]Figure A6 Panel A shows the total number of fires for regulated and unregulated facilities. In general, regulated mills have a higher number of associated harvest fires than unregulated facilities.

[28]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.)

that denotes the number of agricultural fires in non-sugarcane plots associated to mills. Second, I restrict the timeframe of the fires to the months of June through October, outside of the harvesting season.

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 (1), I estimate the following difference-in-differences specification:

$$Y_{ist} = \alpha + \beta_1 D_i \times \mathbf{1}[t \geq 2015] + \gamma_i + \mu_t + \rho_s + \epsilon_{ist} \tag{2}$$

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 the harvest,[29] and ρ_s are week with respect to harvest fixed effects. Figure A7 shows parallel trends for each of the inputs: sugarcane harvested by mechanical and manual cut, total tons harvested, and number of manual workers. Figure A8 shows parallel trends for outputs such as total sugarcane processed and total sugar produced. These two figures show that there are no significant differences between regulated and exempt facilities at the start of the policy in terms of inputs used or total sugarcane harvested or sugar produced.

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 (1):

$$P_{idmt} = \alpha + \beta_1 D_i \times \mathbf{1}[t \geq 2015] + \gamma_i + \mu_t + \rho_m + \lambda W_{id} + \epsilon_{idmt} \tag{3}$$

Where P_{idm} is the ambient pollution concentration of daily $PM_{2.5}$ in $\mu g/m^3$. I obtain two separate versions of equation (3): one for the pollution associated to the fires location and another for the pollution associated to the mills location.[30] Figure A9 Panel A shows

[29] 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.

[30] Important to note, specification (3) differs between mills and fires pollution since the time resolution of the two satellite pollution products is different: pollution associated to fires can be obtained on a daily basis given the larger extent of sugarcane fires while pollution associated to mills can only be obtained on an annual basis given the geographic extend of mills. Therefore, specification (3) in the case of fires would be at the year level instead of the day level.

the parallel trends for the pollution associated with the fields and Figure A9 Panel B shows the pollution parallel trends associated with the mills.

3.2 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 how these emissions are distributed across populations.[31] 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 A10 shows the characteristics of the populations exposed to mills and fields. Generally, poorer households tend to live in rural areas that are exposed to sugarcane fires and mills are located in urban and semiurban areas.

In order to explore the distributional consequences of incomplete regulation, I calculate the catchment areas of all localities (either urban or rural) by creating a buffer of 10 km surrounding the centroid of the community.[32] I then merge these catchment areas to pollution concentrations by predicting the pollution exposure coming from the policy in equation (3), obtaining the annual predicted $PM_{2.5}$ from the policy, $\hat{P}_{jt}$, and modifying the empirical specification of Hernandez-Cortes and Meng (2020):

$$\hat{P}_{jt} = \gamma_0 + \gamma_1 DAC_j \times \mathbf{1}[t \geq 2015] + \tau_j + \mu_y + \epsilon_{jt} \tag{4}$$

[31]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.

[32]I assigned pollution from fires and mills by calculating a receptor catchment area of 10km from the centroid of the urban or rural community. 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.

Where $\hat{P}_{jt}$ is the predicted average annual pollution exposure in community j in year t coming from the policy calculated in equation (3) and DAC is an indicator variable that equals one if the community is disadvantaged (high or very high marginalization index), τ_j are community fixed effects, μ_y are year fixed effects.[33] Standard errors are clustered at the community level. $\gamma_1 > 0$ would imply that disadvantaged communities have experienced a higher burden of the pollution change due to the incomplete regulation compared to other disadvantaged localities whereas $\gamma_1 > 0$ would imply that disadvantaged communities have experienced a lower burden of the pollution change due to the incomplete regulation. I also divide communities into the marginalization categories to examine heterogeneity across different marginalization levels.

3.3 Health impacts of incomplete regultation

This section analyzes whether pollution concentration increases caused by incomplete regulation translate into negative outcomes for populations located within the fires catchment area. In particular, I use the predicted pollution exposure derived in equation (3) to explain changes in birth outcomes:

$$H_{jd} = \alpha + \beta_1 \hat{P}_{j(d-w)m} + \gamma X_{jd} + \lambda_j + \mu_t + \epsilon_{jd} \tag{5}$$

Where H_{jd} denotes average birth outcomes such as birth weight, gestation length, very low birth weight, and very preterm births at the community and day level. $\hat{P}_{j(d-w)m}$ is the predicted exposure coming from the policy calculated in equation (3) associated with the weeks w before the birthdate in the last trimester of the pregnancy, where $w \in \{4, 8, 12\}$. X_{jd} are controls such as average mothers' age and average total doctor visits at the community and day level. μ_t denotes year fixed effects. Standard errors are clustered at the community 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 $\hat{P}_{j(d-w)m}$ is obtained using variation that exploits the introduction of regulation to sugarcane mills with a rich set of

[33]Note that the analysis in equation 4 is done at the *year* level. While the pollution from fires can be analyzed at the day level, I performed the analysis a the year level to have consistency with the pollution from mills, who can only been analyzed at the year level due to the spatial resolution of mills.

controls and fixed effects, specification (5) is likely capturing pollution and not economic activity.

4 Results

4.1 Effects on sugarcane fires

This section discusses the effects of incomplete regulation on within-supply chain leakage. Given that fires are a production substitute for more processing in the plant, we would expect the amount of fires to increase after the boiler regulation in regulated facilities. 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 13.8% after the policy began with respect to the baseline number of fires.

Table 1: Effect of emission limits on daily fires

	(1) Total SHFs	(2) Total SHFs
After 2015 × Regulated	0.044**	0.138**
	(0.022)	(0.066)
Pre 2015 mean	0.297	0.302
Obs.	67,770	66,515
R-squared	0.089	
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mill	Robust
Poisson	No	Yes

Notes: Regulated is an indicator variable that equals one if the mill is regulated by the policy. 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 1. Column (2) estimates the same specification in equation 1 using a Poisson model with robust standard errors. Standard errors clustered at the mill level in parenthesis.

Robustness: Table A2 Column (2) shows the results from equation 1, analogous to column (1) of Table 1 with bootstrap standard errors. Table A2 Column (3) 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, which is consistent with attenuation bias due to measurement error. Table A2 Column (4) and Column (5) show 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. The coefficients remain unchanged but are less significant compared to the benchmark specification. Table A2 Column (6) 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. Table A2 Column (7) and Column (8) show the monthly results controlling for oil and sugar prices, respectively.

Falsification tests: Table A3 Column (1) shows the results for the first falsification test: other agricultural fires in non-sugarcane plots. There is no difference in the number of non-sugarcane agricultural daily fires after the policy for the regulated facilities compared to the unregulated facilities. Table A3 Column (2) restricts the sample of sugarcane fires to the months outside of the sugarcane harvest season (June-October). There is no significant difference in the number of sugarcane fires outside of the sugarcane harvest season either.

4.2 Effects on input substitution

Next, I 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. I find that the amount of sugarcane harvested using mechanical cut decreases although this result is not statistically significant. Given that the use of fires is consistent with an increase in manual cut, I also find that the number of field workers increases by 6%, as Column (3) of Table 2 shows. The results of Table 1 and 2 show that 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 A4 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.[34]

Table 2: Effect of emission limits on weekly input use

	(1) Manual cut (tons)	(2) Mechanical cut (tons)	(3) Total field workers
After 2015 × Regulated	2,921.192**	-278.930	82.954**
	(1,163.016)	(795.990)	(33.433)
Pre 2015 Mean	32,140.435	6,810.680	1,490.736
Obs.	5,887	5,894	5,422
R-squared	0.745	0.727	0.909
Year FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes
Cluster level	Mill	Mill	Mill

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 Equation (2). 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 2. 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 Equiation (2). Standard errors clustered at the mill level in parenthesis.

Although there is an increase in manual cut, I do not find evidence of an increase in total sugar produced. Table A5 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 input reallocation but this does not imply that final output increased.[35] I find suggestive evidence that the non-increase in total sugar produced is due to changes in production efficiency. I estimate changes in production efficiency by using three indicators: (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 A6 shows these results which 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.

[34] 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.

[35] 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, this unlikely affects the identification strategy for two reasons: (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.

4.3 Effects on ambient pollution

What is the effect of supply chain leakage on total pollution concentrations? Table 3 shows the implications of an increase of fires in terms of local ambient pollution levels 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 1.18 $\mu g/m^3$ of $PM_{2.5}$ or a 7% increase in pollution coming from the fields associated to the 2015 regulation. 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 in the areas surrounding the mills.

Table 3: Effect of emission limits on pollution concentrations to nearby communities

	(1)	(2)	(3)	(4)
	Pollution in fields		Pollution in mills	
	$PM_{2.5}$	$\text{Log}(PM_{2.5})$	$PM_{2.5}$	$\text{Log}(PM_{2.5})$
After 2015 × Regulated	1.178**	0.054*	-0.432**	-0.027
	(0.549)	(0.030)	(0.204)	(0.017)
Pre 2015 Mean	17.765	2.621	13.820	2.596
Obs.	20,127	20,127	324	324
R-squared	0.472	0.568	0.926	0.939
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes		
Cluster level	Mill	Mill	Mill	Mill

Notes: Columns (1)-(2) shows changes in pollution concentrations ($PM_{2.5}$ and $\log(PM_{2.5})$) in the sugarcane fields using the difference-in-differences estimator of the impact of the regulation after the policy started following Equation (3) using Buchard et al. (2016). Column (3) and (4) show changes in pollution concentrations ($PM_{2.5}$ and $\log(PM_{2.5})$) in the area surrounding the mills using the difference-in-differences estimator of the impact of the regulation after the policy started following Equation (3) using Hammer et al. (2020). Standard errors clustered at the mill level in parenthesis.

Magnitudes comparison: 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.

4.4 Who experiences the increases in 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. This section uses 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. Table 4 shows that the annual pollution concentrations in disadvantaged communities increased after the policy compared to non-disadvantaged communities. Column (1) shows the average annual predicted concentrations in communities exposed to fires and Column (2) shows the average annual predicted concentrations in communities exposed to mills. A higher amount of communities are exposed to fires and they have higher levels of pollution before the policy. I estimate that by the end of the policy, disadvantaged localities in the fires catchment area experienced 3% more pollution exposure than non-disadvantaged localities relative to the no policy counterfactual. While disadvantaged communities in the mills catchment area also experienced a significant increase in pollution concentrations due to the policy, the effect is very low, a 0.15% increase relative to the no policy counterfactual.

Table 4: Distribution of the average annual effects of pollution from fires and mills

	(1) $\widehat{PM_{2.5}}$	(2) $\widehat{PM_{2.5}}$
After 2015 × DAC	0.438***	0.020***
	(0.042)	(0.006)
Pre 2015 mean	14.826	12.670
Obs.	93,852	17,075
R-squared	0.860	0.681
Year FE	Yes	Yes
Community FE	Yes	Yes
Cluster level	Community	Community

Notes: Columns (1) and (2) show the predicted difference in pollution exposure for disadvantaged communities after the policy. Community exposure was calculated using catchment areas: 10km radii surrounding the community sector. Disadvantaged communities were classified using the 2010 Marginalization Index calculated by CONAPO.

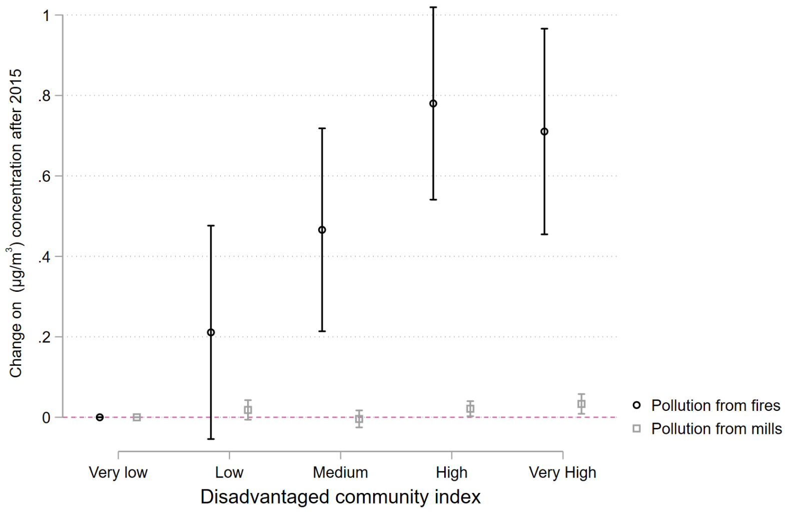
The pollution redistribution impacts are heterogenous with respect to different levels of marginalization: the highest level of marginalization experience a higher burden of pollution exposure than communities with low or very low marginalization indices. This means that more socioeconomically disadvantaged areas experience a higher increase in pollution driven by the policy than other non-disadvantaged areas. Panel a) of Figure 1 shows these results. I also show the robustness of this result using another social vulnerability index, the "Social Lag Index".[36] Figure 1 shows that the exposure from mills did not change 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. 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.

Besides showing that the policy generated a relatively higher pollution exposure for the most vulnerable communities, I test whether locations with higher poverty levels are the most affected by the fires. Table A7 shows that the increase in fires is higher in localities that have poverty levels higher than the 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.

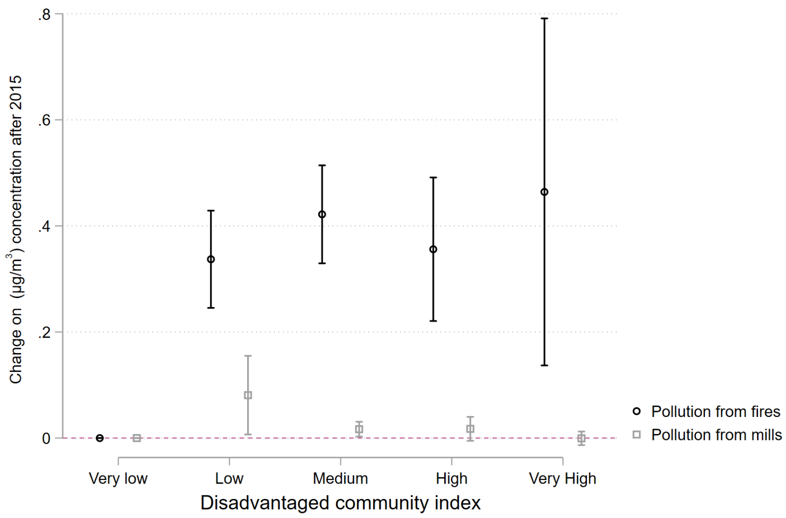
[36]This index considers data from the 2010 data based on different variables than the Marginalization Index calculated by CONAPO. The index considers indicators of infrastructure at the community level and asset holding characteristics for the localities' households.

Figure 1: Distribution of the effects of pollution associated with emission limits

Panel a) Social Vulnerability Index



Panel b) Social Lag Index

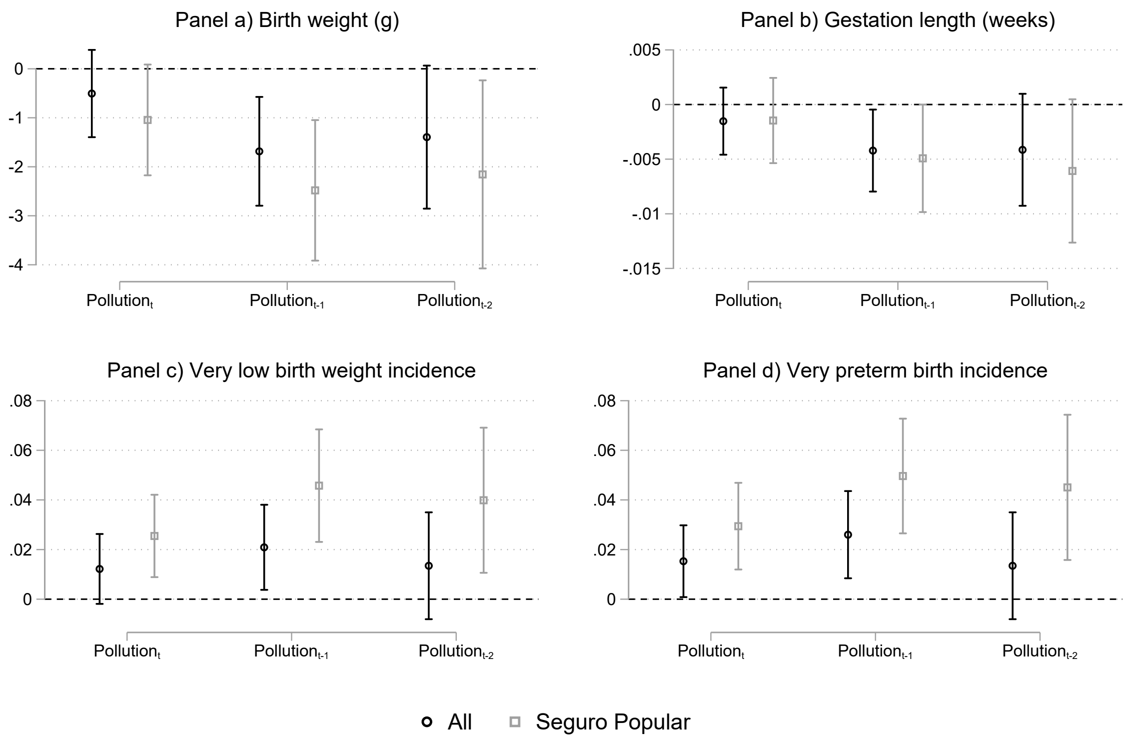


Notes: Panel a) shows the results of pollution associated to fires by level of marginalization according to the marginalization index and the social lag index. Panel b) shows the results of pollution associated to mills by level of marginalization according to the marginalization index and the social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

4.5 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 predicted pollution obtained in Section 4.3 changes birth weight, gestational length, very low birth weight incidence, and very preterm birth incidence for populations located in the fires' catchment area. Figure 2, Table A8, and Table A9 shows the main health results of pollution exposure in the last pregnancy trimester on birth outcomes for all mothers and mothers in Seguro Popular averaged at the community-day level. 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.

Figure 2: Effect of incomplete regulation on birth outcomes



Notes: Panels a)-d) show changes in birth outcomes associated with predicted pollution exposure with 4, 8 or 12 weeks from birthdate for "all" and "Seguro Popular" covered-mothers. Each estimate is obtained from separate regressions. "Very low birth weigh" is defined if weight < 1,500g, "Very preterm birth" is defined if birth length < 32 weeks. All regressions control for average mothers' age and total of doctor visits during pregnancy and year and community fixed effects. "All" includes all mothers, regardless of health care coverage. "Seguro Popular" includes mothers covered by *Seguro Popular*. Confidence intervals calculated using standard error at the community level.

The results in Figure 2 imply that an additional μg/m^3 of PM$_{2.5}$ in weeks 1-8 of the last trimester of pregnancy associates with a birth weight decrease of 1.7 grams on average birth weight across all populations. Rangel and Vogl (2019) 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 on 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, 2019). My estimates are smaller in magnitude than Rangel and Vogl (2019) which could be due to existing differences in fire activity intensity in Brazil, as well as differences in the studied pollutants and identification strategies.[37]

For the other variables analyzed, I find that a unit increase of $PM_{2.5}$ is associated with a 4.2% increase in the very low birth weight prevalence ($<$ 1,500g) and a 5.3% increase in the probability of very preterm birth ($<$ 32 weeks). These results are consistent with other studies finding that fires increase the likelihood of very low birth weight and very preterm birth in exposed populations (Rangel & Vogl, 2018).[38] Worse birth outcomes are found for mothers affiliated to *Seguro Popular*, which is the health system covering the most vulnerable populations in Mexico that are uninsured through the traditional social security networks for formal sector workers (IMSS or ISSSTE). Results in Figure 2 and Table A9 shows that the mothers affiliated to Seguro Popular experience a higher impact of pollution increases: an increase in pollution linked to incomplete regulation increases the very low birth weight incidence by 11.5% and the probability of very preterm birth by 12.4%, double as much as the overall effect for all mothers. These results suggest that increases in pollution due to input substitution are linked to worse health outcomes for populations located in the fires catchment area and these effects are higher for the most socioeconomically vulnerable mothers. These estimates will be used in Section 4.6 to calculate a tax that would internalize the health costs of sugarcane burning cause by incomplete regulation.

4.6 Internalizing the costs of burning

One alternative to alleviate the damages of incomplete regulation is to regulate the pollution from fires with a tax and compensate affected populations. In many settings, a solution to this leakage problem is to extend a Pigouvian tax beyond the jurisdiction of the regulation via a border adjustment that it is output-based (Fowlie et al., 2021). I consider a similar policy that internalizes the existing health costs associated with agricultural burning by calculating an input tax per ton of sugarcane harvested using manual cut levied at the mill level. To do so, I obtain the difference between the social cost per unit of harvested sugarcane and the private cost of production paid by the sugarcane wholesale market and find the tax that would compensate communities affected by sug-

[37]Rangel and Vogl (2019) estimate this by comparing upwind and downwind fires which likely provide a more precise estimate of the impacts of pollution exposure.

[38]Rangel and Vogl (2019) 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.

arcane burning. I then calculate this tax in the context of the sugarcane sector in Mexico using the estimated responses to the incomplete regulation.

Using the parameters obtained in the previous sections and in existing studies summarized in Table A10, 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 \$20 USD (\$403 MXN) per ton of sugarcane, where the average price in 2017 was \$650 USD, meaning is a 3% tax per ton of sugarcane.

5 Conclusion

Many environmental policies are incomplete, covering some sectors or facilities, and not others. This paper shows that under incomplete regulation, regulated facilities can cause increases in pollution in already disadvantaged areas. I find that following the introduction of a regulation aimed at decreasing industrial pollution, regulated sugar mills increased the sugarcane harvest fires and complementary inputs (i.e. manual labor) in their fields. This increase in fires caused an increase in pollution near sugarcane fields. The increase in pollution as a response to the regulation is concerning due to poverty level differences between the populations close to the mills and fields. I find that the pollution increase was higher for disadvantaged communities and translated into worse neonatal outcomes, especially for low-income households. This result contributes to the current discussions on the determinants of environmental justice by examining 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. 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.

This paper also finds that despite increasing pollution in rural areas, input reallocation can have positive employment benefits. Studies have documented the tradeoff between health and local economic outcomes in developing countries (von der Goltz & Barnwal, 2019; Rangel & Vogl, 2019). By showing that 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 to 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.

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 crop burning 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 show that restricting fires to the official distribution areas increases the estimated coefficients. 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. (2013). Leakage, welfare, and cost-effectiveness of carbon policy. *American Economic Review*, *103*(3), 332–37.

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

Becker, R., & Henderson, V. (2000). Effects of air quality regulations on polluting industries. *Journal of political Economy*, *108*(2), 379–421.

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.

Bushnell, J., & Chen, Y. (2012). Allocation and leakage in regional cap-and-trade markets for co2. *Resource and Energy Economics*, *34*(4), 647–668.

Bushnell, J. B., & Mansur, E. T. (2011). Vertical targeting and leakage in carbon policy. *American Economic Review*, *101*(3), 263–67.

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., & Shimshack, J. P. (2020). Environmental disparities in the global south: Evidence from toxic water pollution in urban Mexico.

Chan, N. W., & Morrow, J. W. (2019). Unintended consequences of cap-and-trade? evidence from the regional greenhouse gas initiative. *Energy Economics*, *80*, 411–422.

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.

Copeland, B. R., & Taylor, M. S. (2004). Trade, growth, and the environment. *Journal of Economic literature*, *42*(1), 7–71.

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.

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.

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

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.

Levinson, A., & Taylor, M. S. (2008). Unmasking the pollution haven effect. *International economic review*, *49*(1), 223–254.

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

Perino, G. (2015). Climate campaigns, cap and trade, and carbon leakage: Why trying to reduce your carbon footprint can harm the climate. *Journal of the Association of Environmental and Resource Economists*, *2*(3), 469–495.

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.

Tanaka, S., Teshima, K., & Verhoogen, E. (2021). *North-south displacement effects of environmental regulation: The case of battery recycling* (Tech. Rep.). National Bureau of Economic Research.

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

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:

$$\pi = p[k^\sigma + l^\sigma]^{1/\sigma} - rk - wl \tag{6}$$

The optimal share between l and k is given by:

$$l^* = \left(\frac{w}{r}\right)^{\frac{1}{\sigma}} k^* \tag{7}$$

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 $H_L = \mathcal{H} + \beta M_L(\gamma_l) - D_L(\gamma_l)$.[39] $\mathcal{H}$ 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.[40] 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.[41] 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:

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

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

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:

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

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

[39] 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).

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

[41] 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.

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

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:

$$\pi = p[k^\sigma + l^\sigma]^{1/\sigma} - rk - wl - \tau\gamma(k) \tag{8}$$

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

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

Expression (7) becomes:

$$l^\tau = \left(\frac{w}{r + \tau\gamma_k(k)}\right)^{\frac{1}{\sigma}} k^\tau \tag{9}$$

The optimal demand of l is given by:

$$l^{\tau} = \bar{y}\left[\frac{w^{\frac{\sigma}{\sigma-1}}}{(r+\tau\gamma_k'(k))^{\frac{\sigma}{\sigma-1}} + w^{\frac{\sigma}{\sigma-1}}}\right]^{1/\sigma} \tag{10}$$

Taking the derivative of (10) 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 $\beta(I_R + \phi Y(\tau)) < 2D(\gamma_k)c_M$.

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 $\beta(I_R + \phi Y) < 2D(\gamma_k)$.

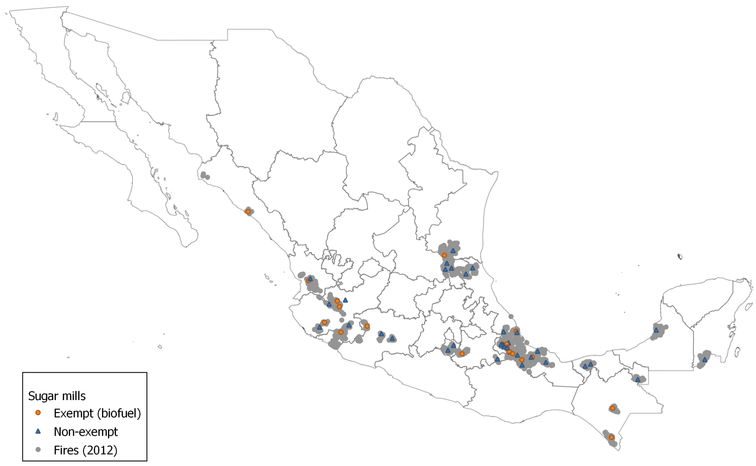
This conceptual framework predicts that under incomplete regulation, the emissions from the regulated technology decrease whereas the use of the unregulated technology and its emissions increase, conditional on being substitutes in the production process. This increase in pollution is regressive, affecting 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.[42]

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.

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

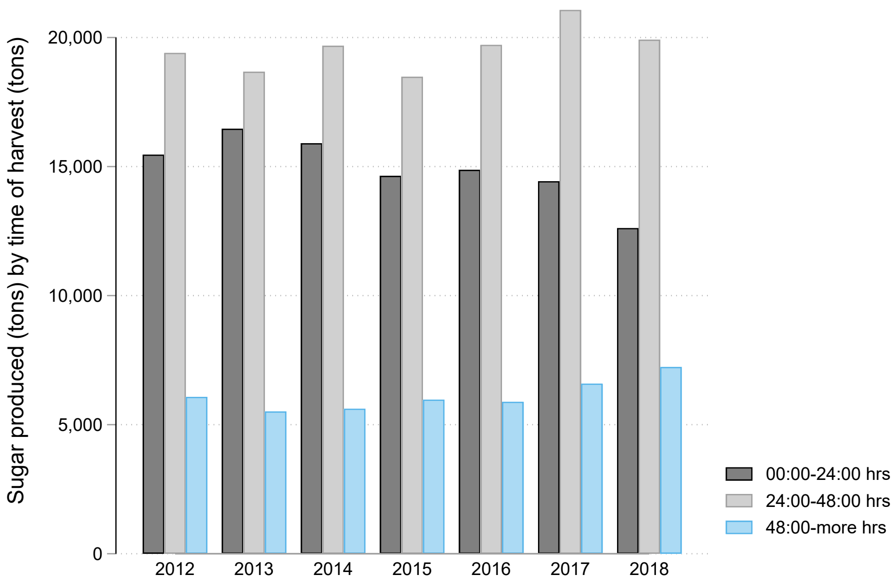
Appendix B: Additional Figures

Figure A1: Mills location



Notes: location of mills classified by regulated and unregulated status by NOM-085-SEMARNAT-2011. Shadow areas are the centroid of the sugarcane fires in 2012.

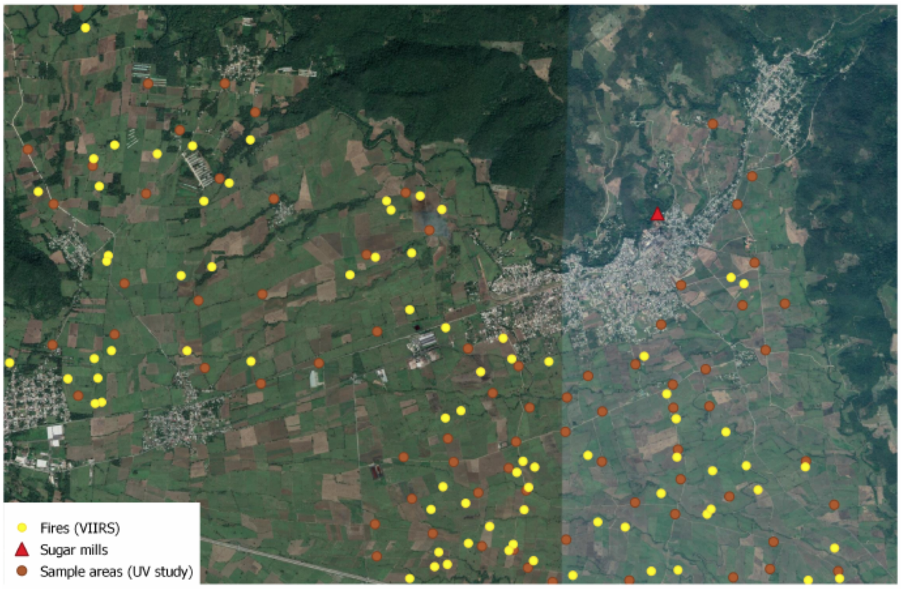
Figure A2: Time to process harvested sugarcane



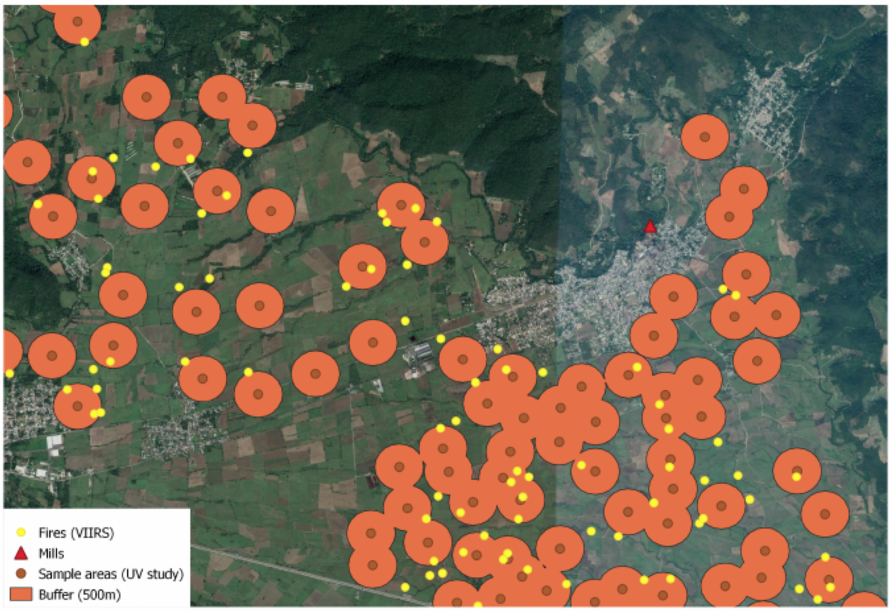
Notes: The figure shows the sugarcane processed (in tons) and its approximate time for processing in the mill after it was harvested by year.

Figure A3: Distribution areas

Panel a)

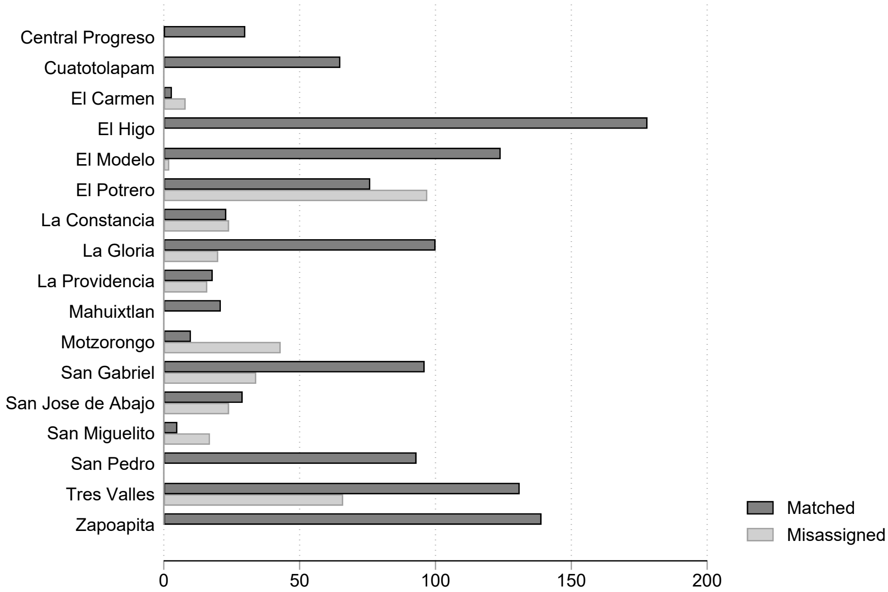


Panel b)



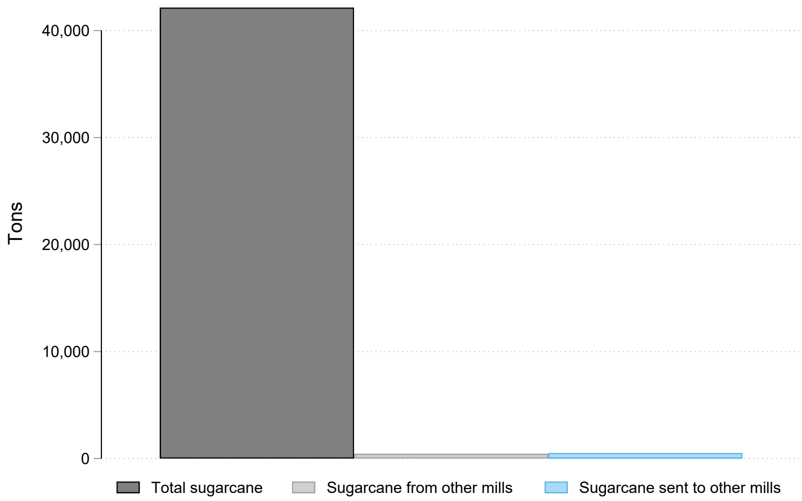
Notes: Panel a) shows the geographic extent of the sample areas (obtained by the Universidad Veracruzana study), sugarcane harvest fires (VIIRS), and a sugar mill. Panel b) shows the geographic extent of a 500m buffer surrounding the sample areas obtained by the Universidad Veracruzana study, along with 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.

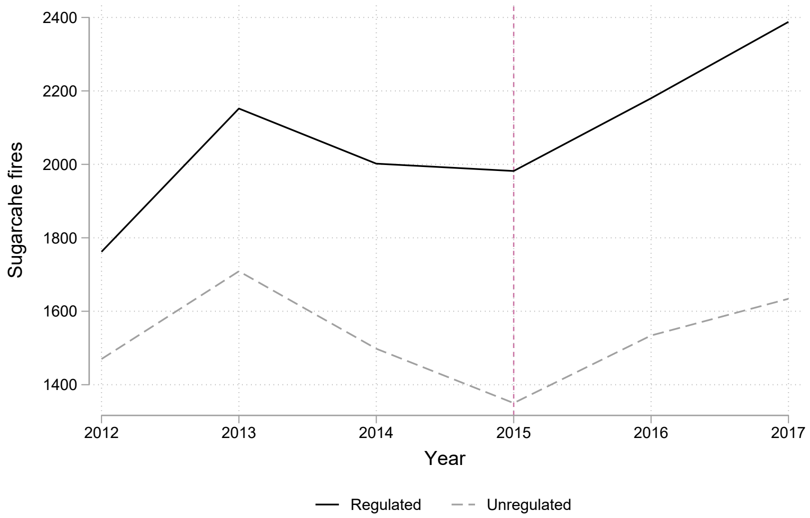
Figure A5: Sugarcane processed by ownership



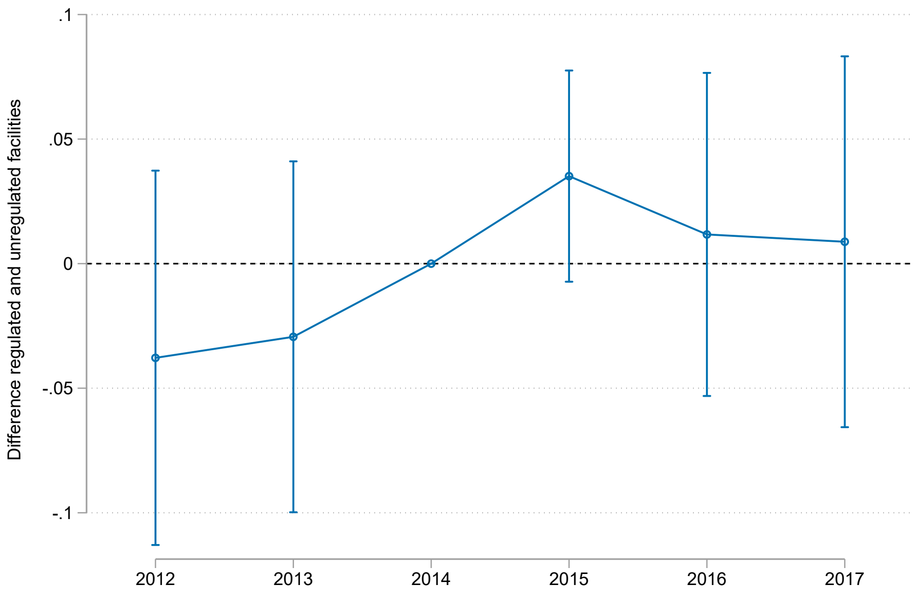
Notes: Total sugarcane processed (in tons) by type of land ownership. Total sugarcane is the sum of sugarcane from own/associated fields, sugarcane from other mills and sugarcane sent to other mills. Sugarcane from other mills is the total sugarcane sent from the fields owned/associated to other mills. 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 A6: Fires associated with regulated and unregulated facilities

(a) Raw mean differences

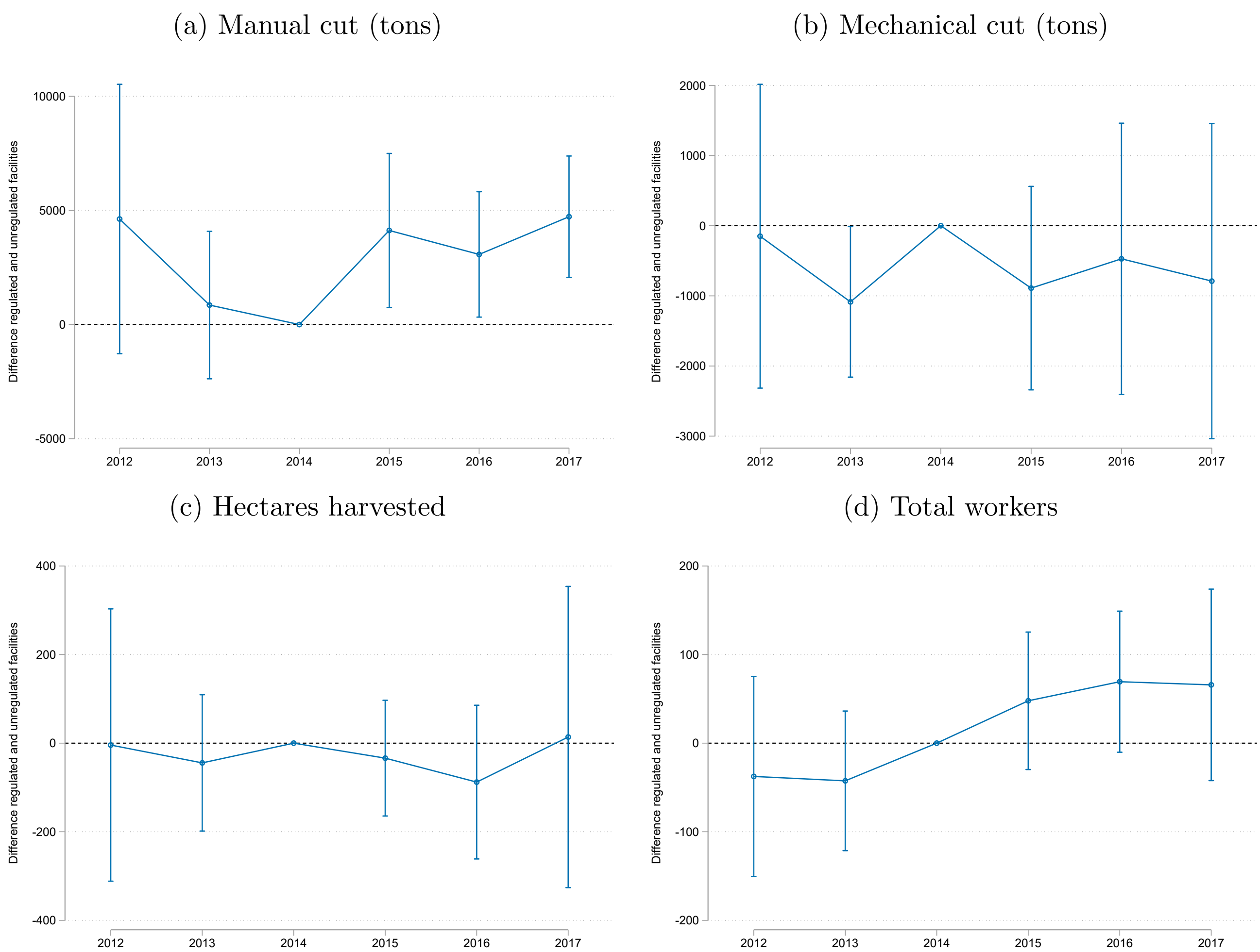


(b) Event-study design



Notes: Panel A shows the mean sugarcane-harvest fires by type of facility. Regulated facilities are defined as facilities using oil as main fuel. Unregulated facilities are defined as facilities using biofuels as main fuel. Vertical axis is the mean fires in sugarcane plots. Panel B shows the difference in differences year specific coefficients for the total number of daily fires following Equation (1). The regulation started in 2015. 95% confidence intervals calculated using cluster standard errors at the mill level.

Figure A7: Inputs-parallel trends



(a) Manual cut (tons)

(b) Mechanical cut (tons)

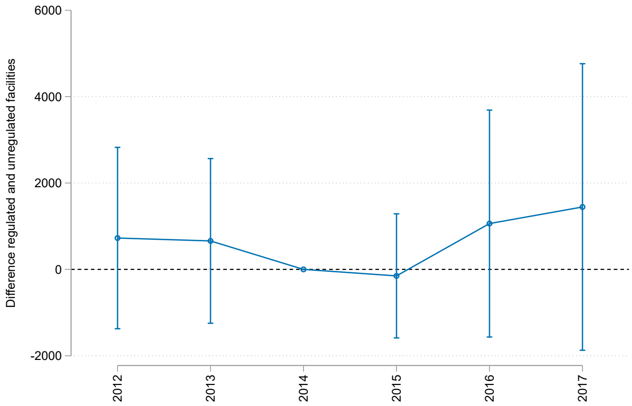
(c) Hectares harvested

(d) Total workers

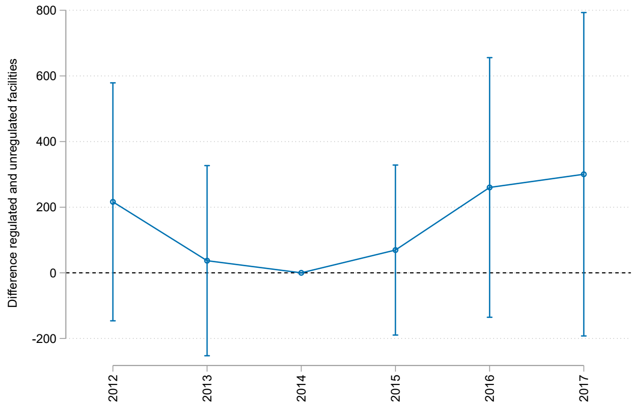
Notes: Panel a) shows the difference-in-differences year specific coefficients for the total sugarcane in tons using manual cut following Equation (2). Panel b) shows the difference-in-differences year specific coefficients for the total sugarcane in tons using mechanical cut following equation Equation (2). Panel c) shows the difference-in-differences year specific coefficients for the harvested hectares following equation Equation (2). Panel d) shows the difference-in-differences year specific coefficients for the total number of manual workers following Equation (2). 95% confidence intervals calculated using cluster standard errors at the mill level.

Figure A8: Outputs-parallel trends

(a) Processed sugarcane



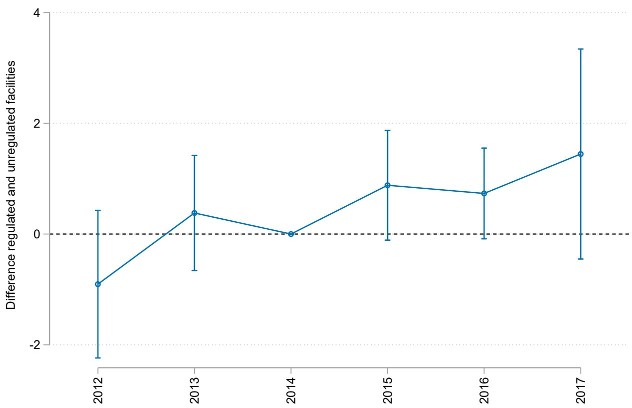
(b) Total sugar produced



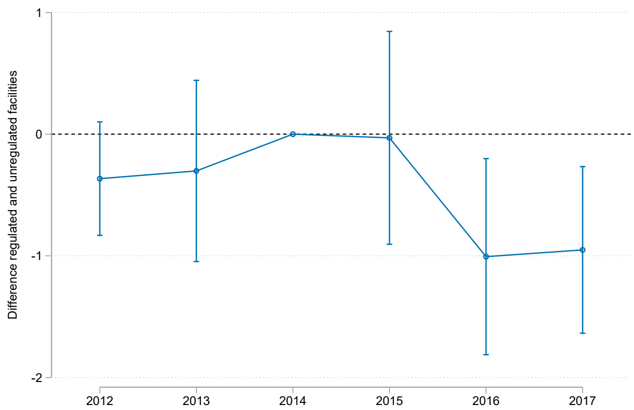
Notes: Panel a) shows the difference-in-differences year specific coefficients for the total sugarcane processed at the mill (mills) following Equation (2). Panel b) shows the differences in differences-year specific coefficients for the total produced sugar at the mill (tons) using mechanical cut following Equation (2). 95% confidence intervals calculated using cluster standard errors at the mill level.

Figure A9: Pollution-parallel trends

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

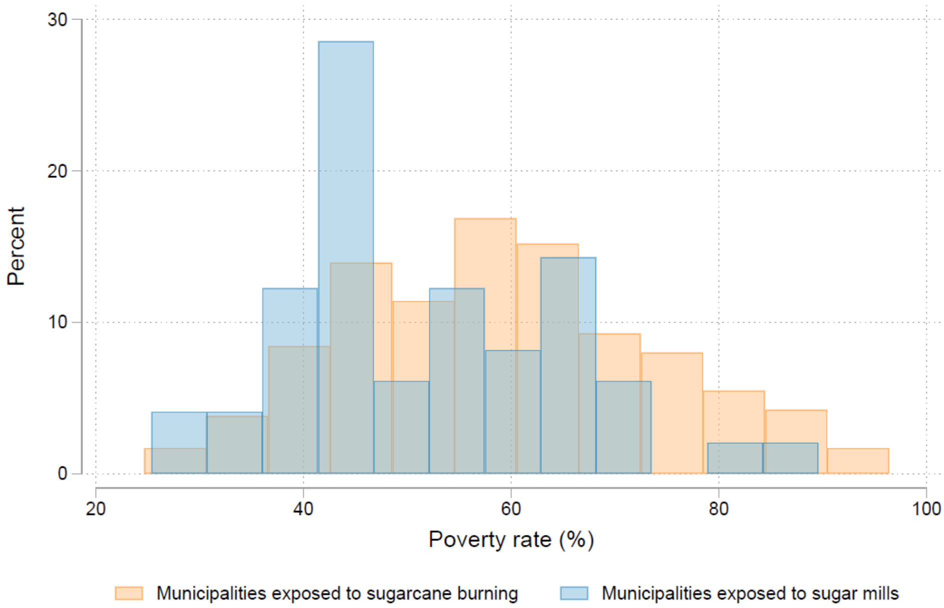


(b) $PM_{2.5}$ pollution ($\mu g/m^3$) in mills



Notes: Panel a) shows the difference-in-differences year specific coefficients for pollution from $PM_{2.5}$ in the fields following Equation (3). Panel b) shows the differences in differences-year specific coefficients for pollution from $PM_{2.5}$ in the mills following Equation (3) using Hammer et al. (2020). 95% confidence intervals calculated using cluster standard errors at the mill level.

Figure A10: 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) Unregulated facilities	(2) Regulated 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)
Total sugar produced per day (t)	555.64	828.56
	(269.05)	(399.47)
Observations	3006	1517

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. 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: Robutness checks

	(1) Benchmark	(2) Bootstrap	(3) Distribution areas	(4) Sample rest. 1	(5) Sample rest. 2	(6) Monthly fires	(7) Sugar price controls	(8) Oil price controls
After 2015 × Regulated	0.044**	0.044***	0.023**	0.043*	0.044*	1.217*	1.217*	1.217*
	(0.022)	(0.014)	(0.011)	(0.022)	(0.023)	(0.653)	(0.653)	(0.652)
Pre 2015 mean	0.297	0.297	0.066	0.296	0.300	8.734	8.734	8.734
Obs.	67,770	67,770	25,000	66,515	66,515	2,268	2,268	2,268
R-squared	0.089	0.089	0.027	0.091	0.089	0.441	0.442	0.447
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mill FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes		Yes	Yes	Yes	Yes	Yes
Cluster level	Mill	Bootstrap	Mill	Mill	Mill	Mill	Mill	Mill
Poisson	No	Yes	No	No	No	No		

Notes: Column (1) shows the baseline identification, 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 (1). Column (2) shows the baseline specification using bootstrapped standard errors. Column (3) shows the baseline specification restricted to the plots that are located in the parcels linked to the distribution network based on 2009 sampling data. Column (4) shows the baseline specification without one mill that invested on a biofuel boiler. Column (5) shows the baseline specificaction without one mill that had the lowest number of matches of parcels. Column (6) shows the baseline specification aggregating at the month level instead of day level. Column (7) shows the baseline specification at the month level controling for international sugar prices. Column (8) shows the baseline specification at the month level controling for Mexican oil prices. All columns use cluster standard errors at the mill level.

Table A3: Falsification tests

	(1) Total NSHFs	(2) Total SHFs
After 2015 × Regulated	0.24180	0.00218
	(0.25468)	(0.00590)
Pre 2015 mean	2.995	0.034
Obs.	71,535	52,326
R-squared	0.321	0.054
Year FE	Yes	Yes
Month FE	Yes	Yes
Mill FE	Yes	Yes
Weather controls	Yes	Yes
Cluster level	Mill	Mill

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 in other agricultural plots without sugarcane using Equation (1). Non-sugarcane fires were classified using land cover data from INEGI. Column (2) shows the difference-in-differences estimator restricting the sample to the months outside of the harvesting season (June-Octuber). Cluster standard errors at the mill level.

Table A4: Effects on wages

	(1) Wage	(2) Wage	(3) Wage	(4) Wage	(5) Wage
After 2015 × Regulated	-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 Equation 2. 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 regulated or all unregulated), therefore deleting 4 mills. Standard errors clustered at the mill level in parenthesis.

Table A5: Effects on outputs

	(1) Raw processed sugarcane (t)	(2) Total sugar produced (t)
After 2015 × Regulated	83.171	78.483
	(869.17986)	(128.52351)
Pre 2015 Mean	41,227.540	4,601.424
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	Mill	Mill

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 Equation (2). 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 Equation (2). Standard errors clustered at the mill level in parenthesis.

Table A6: Effects on efficiency

	(1) Sugarcane processing efficency	(2) Sugar production efficiency	(3) Sugar extraction efficiency
After 2015 × Regulated	-0.857	-4.678	-7.528
	(3.99824)	(5.29413)	(7.27476)
Mean	111.951	99.514	18.005
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	Mill	Mill	Mill

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 2. 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 Equation (2). 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 clustered at the mill level in parenthesis.

Table A7: Effect on sugarcane fires by poverty level

	(1) Total SHFs	(2) Total SHFs	(3) Total SHFs
After 2015 × Regulated	0.04482**	0.02303	0.06633*
	(0.02160)	(0.02415)	(0.03557)
Mean	0.297	0.263	0.339
Obs.	67,770	37,650	30,120
R-squared	0.085	0.073	0.094
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	Mill	Mill	Mill
Poverty level (wrt median)	All	Lower	Higher

Notes: The table shows the results from Equation (1) 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. Standard errors clustered at the mill level in parenthesis.

Table A8: Effect of incomplete regulation on birth outcomes of all mothers

	(1) Birth weight (grams)	(2) Gestation length (weeks)	(3) Very low birth (weight<1,500g)	(4) Very preterm birth (weeks<32)
			Panel a)	
Pollution in w-4 weeks	-0.50539	-0.00152	0.00012*	0.00015**
	[0.45429]	[0.00157]	[0.00007]	[0.00007]
Observations	100,745	100,745	100,745	100,745
			Panel b)	
Pollution in w-8 weeks	-1.68469***	-0.00421**	0.00021**	0.00026***
	[0.56628]	[0.00191]	[0.00009]	[0.00009]
Observations	78,827	78,827	78,827	78,827
			Panel c)	
Pollution in w-12 weeks	-1.39467*	-0.00414	0.00013	0.00020*
	[0.74461]	[0.00261]	[0.00011]	[0.00011]
Observations	61,831	61,831	61,831	61,831
Mean Dep. Variable	3,217.317	38.981	0.005	0.005
R-squared	0.192	0.189	0.172	0.184
Year FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Cluster level	Community	Community	Community	Community

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 community fixed effects. Standard errors clustered at the community level in parenthesis.

Table A9: Effect of incomplete regulation on birth outcomes of low-income mothers

	(1) Birth weight (grams)	(2) Gestation length (weeks)	(3) Very low birth (weight<1,500g)	(4) Very preterm birth (weeks<32)
			Panel a)	
Pollution in w-4 weeks	-1.04404*	-0.00146	0.00025***	0.00029***
	[0.57645]	[0.00199]	[0.00008]	[0.00009]
Observations	61,342	61,342	61,342	61,342
			Panel b)	
Pollution in w-8 weeks	-2.48153***	-0.00492**	0.00046***	0.00050***
	[0.73134]	[0.00251]	[0.00012]	[0.00012]
Observations	47,804	47,804	47,804	47,804
			Panel c)	
Pollution in w-12 weeks	-2.15528**	-0.00608*	0.00040***	0.00045***
	[0.97896]	[0.00335]	[0.00015]	[0.00015]
Observations	36,863	36,863	36,863	36,863
Mean Dep. Variable	3,218.673	39.029	0.004	0.004
R-squared	0.214	0.221	0.174	0.180
Year FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Cluster level	Community	Community	Community	Community

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. Sample restricted to mothers with *Seguro Popular* affiliations. Each estimate is obtained from separate regressions. All regressions control for average mothers' age and total of doctor visits during pregnancy and year and community fixed effects. Standard errors clustered at the community level in parenthesis.

Table A10: Parameters for tax calculation

w	Average ag. wage	\$109 MXN
$\frac{\partial L}{\partial t}$	Table 2, column (1)	9.08%
$\frac{\partial D(\sigma_k)}{\partial \sigma_k}$	Arceo et al. (2016)	0.40%
$\frac{\partial \sigma_k}{\partial k}\frac{\partial k}{\partial t}$	Table 3, column (3)	-3.12%
$\frac{\partial D(\sigma_l)}{\partial \sigma_l}$	Table A8, column (3)	4%
$\frac{\partial \sigma_l}{\partial l}\frac{\partial l}{\partial t}$	Table 3, column (1)	6.63%
r	EPA risk free rate	0.04
$\frac{\partial k}{\partial t}$	Table A6, column (3)	-49%
P	Sugarcane price per ton	\$13,100 MXN
$\frac{\partial y}{\partial t}$	Table A5, column (1)	.20%