Addressing Oil Spills and Agricultural Productivity. Evidence of Pollution in Nigeria.

Beatriz Manotas-Hidalgo
Universidad Pública de Navarra

Working Paper
D.T. 2109

Departamento de Economía
Universidad Pública de Navarra

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Addressing Oil Spills and Agricultural Productivity. Evidence of Pollution in Nigeria. *

Beatriz Manotas-Hidalgo†

September 16, 2021

Abstract

This paper examines how the pollution generated by oil operations in Nigeria can affect agricultural total factor productivity. I analyze oil spills, which are the main ecological disaster in Nigeria and lead to major environmental, economic, and social problems. Following a consumer-producer household framework, and applying a difference-and-difference approach, I estimate an agricultural production function. I find that farmers located less than 10 kilometers from oil spills suffer a relative reduction in agricultural output of around 2.73%. I also examine alternative mechanisms and find that oil-spill pollution can explain my results. I detect less owner-occupied land and a drop in labor income in urban areas close to oil spills, which could also be explained by a decrease in the labor productivity component. This study highlights an externality through which the oil industry affects living conditions in rural areas and stresses the importance of clean-up in areas close to oil spills.

Keywords: Oil Spills, Nigeria, Agricultural Output, Food Security, Natural Resources, Environmental Damages, Conflict.

*I thank Jose Enrique Galdón and Roberto Ezcurra for guidance and support, Miguel Angel Campo-Bescós for guidance in data, Javier Hualde, José Manuel Rodríguez LLanes, Mar Rubio, Valeria Constantini, and Fernando Aragón for helpful comments. This research has benefited from the financial support of the Spanish Ministry of Economy and Competitiveness (Project PGC2018-093542-B-I00).

†INARBE Institute and Department of Economics, Universidad Pública de Navarra. Campus de Arrosadia, 31006 Pamplona, Spain. E-mail address: beatriz.manotas@unavarra.es.

1 Introduction

Food insecurity[1] is driven by multiple factors. Understanding what they are and how they are related to one another is a challenge for scientists working in this field. Specifically, the main variables include conflicts, environmental degradation of livelihoods, climate change, and high volatility in commodity prices. Nigeria is Africaâs most populous country and its largest oil producer. It is a particularly suitable example for studying the links between some of these variables and indicators. Nigeria is a country cursed by natural resources (Sala-i-Matin and Subramaian, 2013), which suffers from complex political issues including endemic corruption, inequality within and between ethnic groups, national disunity, oil disputes, environmental degradation, instability, and poverty. It also faces three sources of violence: Boko Haram insurgency, Middle-belt conflict, and the Niger Delta conflict.

Onshore oil operations are a key aspect related to environmental degradation. They have damaged local soil and water resources, leading to problems in public health in nearby locations (Bruederle and Hodler, 2019). When focusing on the specific effects of oil spills in Nigeria it is useful to analyze negative externalities generated by extractive industries on places distant from the sources of production, e.g. pipeline networks, where traditional agricultural activities are the predominant source of subsistence. Oil spills are the biggest environmental disaster in Nigeria and have exacerbated environmental, social, and economic problems (Madu et al., 2018; Nwankwo, 2015).

The 2011 United Nations Environment Programme report on oil spills in Ogoniland,[2] a region which covers close to 1,000 square kilometers in Rivers State, southern Nigeria, is a turning point that further emphasizes the establishment of Corporate Social Responsibility (CRS) to clean up the area. However, host communities argue a lack of responsibility for environmental damage in the government of the federation and multinational oil corporations. The study also reports that oil spills could affect more territory than the

[1] "Food security exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life. Household food security is the application of this concept to the family level, with individuals within households as the focus of concern. Food insecurity exists when people do not have adequate physical, social or economic access to food as defined above. "The State of Food Insecurity in the World, pg.4. FAO, 2010.

[2] "While oil exploration and production in the Niger Delta began in the late 1950s, operations were suspended in Ogoniland in the early 1990s due to disruptions from local public unrest. The oilfields and installations have since largely remained dormant. However, major oil pipelines still cross through Ogoniland, and oil spills continue to affect the region, due to such factors as a lack of maintenance and vandalism to oil infrastructure and facilities." United Nation Environmental Programme". https://www.unenvironment.org/explore-topics/disasters-conflicts/where-we-work/nigeria/ogonilands-oil-history

areas estimated to be directly affected, through rivers and water bodies, and that effects are long-lasting. Thus, the damage to the environment could be greater than that directly calculated.

Keeping in mind the above conclusions, in this paper I attempt to assess an externality effect of onshore oil operations specifically related to pipelines on the agricultural sector. In particular, I study whether oil spillage shocks are associated with an economically significant reduction in agricultural total factor productivity, and hence, in agricultural production in nearby locations. The assumption is that the oil spills analyzed may not have affected farms directly, but that their could be affected indirectly by the filtration through the soil of nearby contaminated water and by air pollution from fires around the spills. I further hypothesize that the impact could be long-lasting. More specifically, I pose the following questions: Does the presence of oil spills lead to a reduction in agricultural total factor productivity among farmers in nearby locations? Does the effect persist beyond the periods when oil spills happen?

Unlike previous literature that has studied the impact of oil spills on farming in Nigeria, I focus on all regions of the country, using geospatial data from oil spills, The Nigerian Oil Spill Monitor, which provides data collected by the National Oil Spill Detection and Response Agency (NOSDRA), and geo-referenced micro-data from farming households drawn from the Nigeria General Household Survey (GHS-Panel). The former gives information on the locations and dates of around 12,000 oil spills from January 2006 to December 2018. I have also calculated the geographical coordinates of oil spills not reported, dating from before 2013. The latter provides information on agricultural production and agricultural practices from around 1,425 farmers in four waves, covering the harvesting periods from 2010 to 2018. To the best of my knowledge, this is the first paper to use such data to assess the impact of oil spills on agricultural output through agricultural total factor productivity.[3]

In my identification strategy, I first observe the effect of oil-spill pollution on agricultural total factor productivity by estimating an agricultural production function. I use the analytical framework of consumer-producer household models with incomplete markets (Benjamin, 1992; Aragón and Rud, 2016). In these models, production and consumption

[3]This dataset has only been used before by Bruederle and Hodler (2019), who found clear evidence for harmful effects of nearby oil spills on surviving children.

decisions are not entirely separable, and household endowments could be used as inputs. The model helps to determine whether a variation a change in total factor productivity brings about a change in agricultural output, that is whether oil-spill pollution could affect the quality of essential inputs or whether the variation is the result of a change in input uses. Taking this approach that identifies how farmers act, I may be able to distinguish the channels through which pollution from oil spills affects agricultural output.

Second, I consider the empirical challenge posed by the fact that agricultural output could be regularly different in areas where oil spills happen. To overcome this issue, I also explore the methodology proposed by Fenske and Zurimendi (2017) and Aragón and Rud (2016), using a difference-in-difference approach. With this technique, I explore two sources of variation: A proxy of the quantity of pollution caused by spills that could also be persistent over time, and the distance of households from oil spills. This identification strategy means that in the absence of oil spills, any changes in agricultural output should be similar in both areas over time.[4]

As a proxy of accumulated oil spill pollutants, I create a function that covers all oil spill events per location in all the said periods. A key point in that function is that oil spills follow an exponential decay pattern on cultivable land. The same conclusion can be extended to labor productivity and crop yields. I also add further functions to check the robustness of my results.

I find evidence of a significant reduction in both total factor productivity and agricultural output attributed to oil spills. My estimations suggest that an increase of one standard deviation in my measurement of cumulative oil spill pollution is associated with a drop of around 8% in agricultural output in locations within 5 kilometers of oil spills. However, the data also suggests significant effects in areas 7.5 kilometers from oil spills. The results are similar if partial measurements such as crop yields are used. The findings are also robust to different model specifications, e.g. the inclusion of a definition of the proxy of oil-spill pollution based on different persistent effects, quantity of oil lost, additional agricultural practices, soil characteristics, climate variables, and heterogeneous location trends. The consumer-producer framework means that a reduction in agricultural output directly affects the consumption potential of households. Indeed, Oshienemen et

[4]In particular, Bruederle and Hodler (2019) also used locations nearby oil spills as a part of their identification strategy.

al. (2018) report an increase in poverty as an indirect effect in villages near oil spills.

I also explore alternative channels that could explain my results. For instance, following Bruederle and Hodler (2019), I investigate whether the results only affect the Niger Delta area, where there are events associated with oil operation around the extraction sites in these regions, or whether the findings are the consequence of violent attacks on oil infrastructures and other conflict incidents. I further focus on differences in the characteristics of agricultural workers and changes in property rights. Empirical evidence shows that households in locations near oil spills own less land. This result could reflect the risk of land expropriation by the state. For example, oil companies could require land close to pipelines to build access infrastructures. Thus, farmers could invest less in such land, thereby further decreasing agricultural factor productivity. Finally, I find a decline in labor incomes in urban areas close to oil spills, suggesting a drop in labor productivity as a plausible mechanism. These results could support the notion that a reduction in agricultural total factor productivity reflects the well documented cyclical penury and poverty among host communities (Nriagu, 2019; Madu et al., 2018).

To the best of my knowledge, Akpokodje and Salau (2015), Ojimba (2011), Ojimba(2012), and Inoni et al. (2005) also consider the economic effect of oil pollution[5] on crop production in the Niger Delta. Akpokodje and Salau (2015) assess the consequences of oil spills as a catalyst in accelerating deforestation and, hence, indirectly reducing agricultural output. Ojimba (2011) focuses on the economic effects of oil spills on crops, farms, and the size of farmland, while Ojimba (2012) examines the impact of crude oil and gas pollution on crop production. Finally, the empirical evidence provided by Inoni et al. (2005) focuses on the effect of oil spills on crop yield, land productivity, and farm income. All these authors find a negative impact of oil pollution in their main dependent variables. However, in Akpokodje and Salau (2015), the oil spillage variable is not significant, and it acts indirectly, leading to a loss of forest mass

The above results are potentially noteworthy. However, my work differs methodologically from their in several ways. Ojimba (2011) splits the dataset between oil-polluted and non-polluted farms, whereas Inoni et al. (2005) focus on the presence of oil spills with a dummy variable in the harvesting season. I consider that both the number

[5] 5Oil and gas pollution could be driven by the pollution caused by all operations related to the extraction, production, and transport of oil and gas. Like Inoni et al. (2005), I only consider the effect of oil-spill pollution on agricultural output.

and the persistence of oil spills need to be included to assess the effect of oil-spill pollution on agricultural output. Moreover, all the papers mentioned above consider the impact of oil pollution only in particular regions of the Niger Delta,[6] while my research looks at the whole country, considering onshore oil spills far from oil-producing sites. Thus, I may find evidence that the consequences of oil spills in areas close to the oil transport network are at least as detrimental as in locations close to oil wells and gas flares themselves.

Akpokodje and Salau (2015) use country-level data in their analysis, while Ojimba (2011), Ojimba (2012), and Inoni et al. (2005) use data collected from interviews with almost 290 farmers in different locations from of the same regions. The methodology used in these last papers thus means that it is unlikely that their results could be extrapolated to other areas. Geo-referenced data also makes it possible to consider whether the consequences of oil spills can spread to nearby locations, which is a step forward towards assessing the environmental impact of oil spills.

Earlier studies do not adequately address the issue of potential endogeneity, which is particularly important in establishing a causal link between oil spills and agricultural output. In this paper, I use three strategies to tackle this issue. The first is to use district and time fixed effects to control for time-invariant factors affecting both agricultural output and oil spills, such as geographical and seasonal features. The earlier papers referenced do not use any fixed effects. I also use the above-mentioned difference-and-difference approach with the geo-referenced dataset of oil spills to create an oil-spill area around the pipelines and thus control for the issue of omitted variables. Finally, I use instrumental variables to control for the endogeneity of inputs in estimating agricultural production.

At a more general level, my paper contributes to the emerging body of literature at the intersection of environmental economics and development economics. This literature is filling an important gap given that most studies on the economic effects of pollution have hitherto been conducted in developed countries. For instance, the main focus has been to assess the effect of pollution on labor productivity (Graff Zivin and Neidell, 2012), labor supply (Hanna and Oliva, 2011), and human capital accumulation (Currie et al., 2009). Like me, Bruederle and Hodler (2019), Aragón and Rud (2016) and Jayachandran (2009) also focus on the impact of pollution on the extraction of natural resources in developing

[6] Ojimba focuses on Rivers State, while Inoni et al focus on Delta State. Both regions are considered part of the Niger Delta area, which is formed by nine regions, stretching over the Delta of the River Niger, the biggest river in West Africa.

countries. In particular, Bruederle and Hodler (2019) study the causal effects of oil spills on infant mortality in Nigeria, providing some evidence for negative health effects of nearby oil spills on children. Aragón and Rud (2016) provide evidence that the expansion of large-scale gold mining in Ghana lowers total agricultural productivity in places within twenty kilometers of the mines. In contrast with this last paper, I assess the consequences of unexpected events related to the transport of oil on agricultural output, considering that the effects do not disappear in a single period. Jayanchadran (2009) also investigates forest fires originated by palm oil producers and logging companies which burn out of control and affect all of Indonesia. She finds a strong link between air pollution from forest fires and infant mortality. My paper differs from the above (except for the paper by Bruederle and Hodler (2019)) in that I trust in the plausibly random timing of oil spills at locations that were affected at some point during the period from January 2006 to December 2018.

The rest of the paper is organized as follows. Section 2 reviews the background to oil spills and pollution in Nigeria, and their links with agricultural output, and presents the model related to this framework. Section 3 describes the methods, covering the data and the empirical strategy. Section 4 presents the main results and several robustness checks. Section 5 shows alternative specifications, and Section 6 deals with mechanisms. Section 7 concludes.

2 Background

2.1 Oil Spill and Pollution

'Oil spill" means any spill of crude oil or distilled products such as diesel or jet fuel, gasoline, kerosene, Stoddard solvent, hydraulic oils, hydraulic oils, and lubricating oils.[7] "Oil-spill pollution" means the negative effect of oil spills on the environment and living organisms. When an oil-spill event occurs, location is an important predictor of its impact. Onshore spills close to human populations have a greater economic impact. The spillage rate and the number of oil leaks are also decisive determinants of the severity of

[7] "Crude oil and its derivatives include various individual hydrocarbons. Hydrocarbons are constituted "[...] from carbon and hydrogen atoms that bind together in various ways, resulting in paraffins (or normal alkanes), isoparaffins (isoalkanes), aromatics (such as benzene or various PAHs), cycloalkanes and unsaturated alkanes (alkenes and alkynes)â https://www.environmentalpollutioncenters.org/oil-spill/. Other components are sulfur, nitrogen, and/or oxygen atoms.

the consequences (Chang, et al. 2014).

The effect of crude oil pollution on wetland soil, which is what most of the Niger Delta area comprises, is to lower soil fertility by increasing soil PH up to 80%, thus reducing available phosphorus (AP). These effects can alkalinize marsh soil, affecting soil fertility and causing deterioration on wetlands (Wang, et al. 2013). Oil spills often also lead to fires, which release respirable particulate matter (PM from now on) into the air (Bruederle and Hodler, 2019). These air pollutants can be carried over long distances and deposited on the ground as acid rain, or directly absorbed by plants (Aragon and Rud, 2016).

2.2 Oil Spills in Nigeria and Agricultural Output

My empirical analysis deals with pollution from oil spills in Nigeria. Nigeria is a West African country in the Gulf of Guinea with a surface area of 923,773 km2 (Zabbey et al, 2017) and a population estimated at close to 201 million in 2019. The Niger Delta area in Nigeria comprises diverse ecosystems of large forests, freshwater, mangrove, and swamps, which are characterized by continual salt-water-inundations. It is the largest wetland in Africa (Okonofua, 2011).[8] The Niger Delta basin has been studied in depth because of its vast deposits of petroleum resources. Oil operations were started there in the 1930s by the Royal/Dutch-Shell Company, operating under the name Shell Petroleum Development Company (SPDC) (Madu et al, 2018). The first oil was produced in December 1957, and the petroleum sector shaped the Nigerian economy in the early 70s, leading to a rapid accumulation of capital, declining total factor productivity, and contracting utilization of capacity.

The increasing dependence of the Nigerian economy on hydrocarbon extraction has placed severe pressure on components of the environment as a result of incidental and accidental discharges of hydrocarbon components into the environment. “[..]The oil companies operate over 5,284 oil wells and thousands of miles of oil pipelines networks though the Niger Delta region” (Madu et al. 2018, pg. 79). The main environmental challenge is that of oil spills. This is a common issue in many developing oil-producing countries. However, in most cases spills are associated with operational or mechanism

[8] According to the United Nations Development Program Report (UNDP, 2006), most of the people in the area depend on the natural environment for their livelihood. Good agricultural lands, fisheries, and well-developed industries are part of the abundant resources in the region.

failures,[9] but in Nigeria they are also the result of oil theft, sabotage, and pipeline vandalism. [10]

The 2011 United National Environmental Program (UNEP) report looks in depth at the consequences of oil spills in Ogoniland, (River State, Niger Delta), finding "[..]oil contamination severely impacting many components of the environmentâ. In addition, Ogoniland frequently has high rainfall, and when oil spills are not properly cleaned up, oil has been found "[..] being washed away, traversing farmland and almost always ending up in the creeks" (UNEP report, 2011, p. 9). The report also focuses attention on the importance of land/resource use policies in the Niger Delta, and the importance of corporate social responsibility programs, including clean-up programs in the area.

Agriculture has traditionally been the dominant economic activity in Nigeria. In 1985, crop farming and fishing accounted for approximately 90% of all activity in the Niger Delta area. The active labor force linked to these activities accounted for around 50%-68% of the total. More than 90% of farmers are subsistence farmers, working with traditional techniques and basic tools. Land is still farmed using the bush fallow system or land rotation. These organic farming techniques are very susceptible to environmental changes that affect water or soil, and therefore lead to deforestation. This high level of resource utilization based on land and labor-intensive methods makes the area more susceptible to oil pollution.

I link oil pollution with agricultural productivity here through the impact of pollutants on crop yields and health, soil quality, and human capital. Once crude oil and petroleum products leak directly into the environment, different compounds are absorbed by the soil, entering ground and surface water or evaporating in the air depending on their physical characteristics (Bruederle and Hodler, 2019). These pollutants lead to a rapid deterioration in the soil, a reduction in crop yield and, hence, to a fall in agricultural output. Evidence from biological science (for example Maggs et al, 1995; Marshall et al, 1997 between others)

[9] Many pipelines are old and subject to corrosion. The estimated safe life span of a pipeline is fifteen years, but in numerous places in the Niger Delta it is possible to find pipelines 20 or 25 years old. These pipelines are thus prone to rupture and are major fire hazards (Nriagu, 2019).

[10] There are varying socio-political factors related to pipeline vandalism. One important problem associated with unrest in the Niger Delta which results in the destruction of oil pipelines, and consequently oil spills in the area, is who controls oil revenues. Since 1999, the federal government has paid out 13% of the revenues derived from oil to oil-producing states. The federal government uses a revenue allocation formula to distribute these tax revenues to states. The quest for self-determination of young people in the Niger Delta area and a failure to consider the interests of host communities are associated with the control of these rents. This has led to an increase in civil unrest and often to the sabotaging of oil pipelines, thus causing oil spills in the area. (Madu et al. 2018).

finds a "[..] reduction of around 20-60% in the yield of crops such as rice, wheat, and beans. In particular, a case study of the effect of oil pollution on soil properties and growth of tree crops[11] showed that seedling germination and plant heights are significantly affected at high levels of pollution (Uquetan, U. et al. 2017). Besides, the effects on ground pollution could be cumulative and long-lived. In fact, although the Nigerian crude oil has rapid evaporation loss of around 50%, a study carried out about nineteen years after a major 1970 oil spill at Ebubu, Ogoni, found that "[...]vegetation in areas downstream of the spill was still being degraded due to a slow seepage of crude oil from the spill site" (Nriagu, 2019, pg. 761). Consequently, agricultural productivity could be affected not just in the period when the spill occurs but also via a persistent effect that can continue to impact the agricultural cycle years after the events.

Finally, agricultural output could also fall because of a drop in labor productivity. There may be direct adverse health effects on workers through PM inhalation or indirect effects through damage to livelihood resources, such as the quality of foods from degraded lands and fishing grounds. In any case, a drop in labor productivity or labor supply could result, leading to a decrease in agricultural output.

2.3 Analytical Framework

This subsection develops a simple framework for understanding how oil-spill pollution could result in adjustments in the optimal behavior of households. I follow the simple agricultural standard model of consumer-producer households used in development literature as per Aragón and Rud (2016), who extend the framework set out by Benjamin (1992). Oil spills can impact directly on land plots, or may have occurred some periods before and soil could receive filtration that still affects agricultural total factor productivity. Also, there may be no direct impact on plots, but indirect pollution through air or water pollution that leads to changes in soil quality, biomass, and human health.

I assume that farmers (households) are both consumers and producers of an agricultural good with a price of $p = 1$. Households have a productive parameter A and use land, X, and labor, L to produce the agricultural good $Y = F(A, X, L)$. F is a concave production function. Farmers have certain endowments (E^X, E^L), which represent land and household

[11] Specifically, cocoa, cashew, pawpaw and mango.

endowments, respectively. Endowments are used as inputs on the farm or can be sold at a local input market (X^s, L^s), as land and labor supply at prices r and w. Labor endowments also can be used as leisure. In addition, farmers can buy an additional quantity of land and labor (hired labor) when there are producers in line with land and labor demand: (X^b, L^b).

The problem of farmers consists of maximizing household utility $U(C, l)$ over consumption, C, and leisure, l, subject to the budget constraint $C = F(A, X, L) - r(X^b - X^s) - w(L^b - L^s)$. Endowment constraints are $X = E^X + X^b - X^s$, and $L = E^L + L^b - L^s - l$.

In the context of the Nigerian agricultural market,[12] I assume that households are not homogeneous in their access to input markets. [13] In particular, there are two types of farmer: Unconstrained farmers, who participate in competitive input markets, and fully constrained farmers, who neither buy nor sell inputs. In the first case, if input markets exist and work well it is possible to study production and consumption decisions separately, and there is trade. Households maximize their profits and, given the optimal profit, choose between consumption and leisure levels. Thus, the optimal levels of output and inputs, $Y^*(A, w, r)$, $X^*(A, w, r)$, and $L^*(A, w, r)$, depend only on the value of A, which means total factor productivity, and on input prices.

For fully constrained farmers, endowments shape the optimal decisions on inputs. Farmers use all their land in the planting season $X^* = E^X$ given that the opportunity cost of land is zero. Given that there is no labor market, $L^s = L^b = 0$, and the optimal level of labor depends on a trade-off between income and leisure. In this simplified framework, the farmer's problem is:

$$MaxU(C, l)$$

[12] Even though agriculture is the main system of livelihood for Nigerian, the sector is characterized by poor access to input markets. For example, "[..] an outdated land tenure system that constrains access to land (1.8 ha/farming household), a very low level of irrigation development (less than 1 percent of cropped land under irrigation), limited adoption of research findings and technologies, high cost of farm inputs, poor access to credit, inefficient fertilizer procurement and distribution, inadequate storage facilities and poor access to markets have all combined to keep agricultural productivity low (average of 1.2 metric tons of cereals/ha) with high post-harvest losses and waste" (FAO in Nigeria `http://www.fao.org/nigeria/fao-in-nigeria/nigeria-at-a-glance/en/)`

[13] Land acquisition is bound by the Land Use Act of 1978. There are three types of land market: 1) Formal land markets, where the government allocates a certificate of occupancy; 2) combined formal and informal markets where there is a certificate of occupancy in the transfer of land rights; and 3) informal markets. The titles owned do not entail a certificate of occupancy because the bulk of the transactions are not documented. Nevertheless, an estimated 95% of agricultural land in Nigeria is not titled (Oluwatayo et al., 2019).

$$\text{s.a}$$
$$C = F(A, E^X, L)$$
$$L = E^L - l \tag{1}$$

The first order condition becomes: $U_c F_L = U_l$. Thus, the optimal level of labor is a function that depends on both the level of total factor productivity and input endowments, $L^*(A, E^X, E^L)$.[14]

If oil spills directly impact plots in the growing season, the agricultural good that has been planted is lost. In terms of the model, this can be interpreted as a reduction of land endowment, leading to a diminishing land supply. However, there are two indirect channels through which oil-spill pollution affects agricultural output and hence household consumption. As described above, most oil spills are located in the Niger Delta, and are most likely to occur in the pipeline networks close to oil wells. Thus, oil companies that operate in this zone could have a demand for local inputs (land and labor), leading to an increase in input prices, which would reduce input use and consequently, agricultural output among unconstrained farmers. This channel also could reduce the supply of inputs through government expropriation of land for oil extraction and infrastructure access, and through population displacement. There would be no effect on total factor productivity A.[15]

Moreover, oil pollution affects the quality of inputs, as discussed in the previous section, and soil quality, crop yields, health, and labor productivity all decrease. This argument is linked to a drop in the total productivity factor, which unambiguously causes a decrease in agricultural output and household consumption, although input uses may not change. It might also lead to a reduction in input uses. For unconstrained farmers, this might mean a reduction in labor and land uses because input prices do not change. Among constrained farmers, the drop in total factor productivity leads to labor being replaced by leisure, while the use of land does not change.

In short, this model highlights the importance of studying the impact of oil-spill pollution through its indirect effects on agricultural total factor productivity. Other

[14] Agricultural employment for a wage is relatively infrequent in Nigeria. GHS-Nigerian data shows that only 3.5% of men and 1.4% of women are wage workers.

[15] I also explore the use of other inputs such as fertilizers, improved seeds, etc, later in my analysis.

outcomes, such us input uses and agricultural output, may not be very informative about the channels that determine a drop in agricultural output.

However, the unobservable heterogeneity in A could also impact input uses and the engagement of the econometric identification of total factor productivity. Thus, in my empirical approach, I follow the model prediction that relies on household endowments as a key for determining input uses in the presence of imperfect input markets. This assumption leads to consistent estimates of the parameters of the production function

3 Methods

3.1 Data

I merge geo-referenced household surveys containing agricultural, socioeconomic, and weather variables from the Nigerian General Household Survey (GHS-Panel) (waves 1, 2, 3, and 4) with data from oil spills also geo-referenced from The Nigerian Oil Spill Monitor[16] to construct a final dataset of around 6,000 observations for my main analysis.

Agricultural output and inputs

My main data source is a repeated cross-section from the Nigeria General Household Survey Panel (GHS-Panel).[17] It is collected by the National Bureau of Statistics in collaboration with the World Bankâs program on Living Standards Measurement Surveys - Integrated Surveys for Agriculture (LSMS-ISA). This program was revised in 2010 to include a panel component (GHS-Panel). The GHS-Panel is a national survey of 5,000 households, which are also representative of geopolitical zones (at urban and rural levels). Households were interviewed in 2010-2011 (Wave 1), 2012-2013 (Wave 2), 2015-2016 (Wave 3), and 2018-2019 (Wave 4). The Nigeria GHS-Panel is part of a larger, regional project in Sub-Saharan Africa that involves eight countries and seeks to obtain better agricultural statistics. The surveys collect data on agricultural activities, other household income activities, household expenditure, and consumption. The finest level is that of enumeration areas (EA), which approximately match neighborhoods (urban areas) and villages (rural areas). In Wave 4, the GHS-sample was partially refreshed to maintain

[16] https://oilspillmonitor.ng

[17] I cannot estimate a panel dataset because at the time of writing the longitudinal weights were still being prepared. However, cross-section weights are available for each round.

the representativeness and integrity of the sample. A new set of 360 random enumeration areas were incorporated into the sample, which meant 3600 new households. Thus, a subsample of 1425 families from previous rounds was interviewed. Farmers are located in 423 local government areas (LGAs), 37 states, and six Zones: North-Central, North-East, North-West, South-East, South-West, and South-South. Figure B.2 in the Appendix shows a map of the six geographical zones of Nigeria.

Each wave consists of two visits to each household:[18] A post-planting visit just after the planting period to collect information on inputs used, planting preparation, labor used for planting, and other information relating to the period; and a post-harvest visit after the harvest season to collect information on crops harvested, labor used for harvesting, and other variables related to the harvest cycle. I focus on the farming household as a production unit in a period (the year) that represents a season-round pair. A farmer may operate one or more plots of land, so I aggregate any information at the plot level to household level. The GHS panel also provides a set of geospatial variables using household locations and geo-referenced plots together with various geospatial databases that are available to the survey team. Specifically, the geo-coordinates of clusters (or an average of household GPS locations by EA in GHS-Panel) are reported but slightly displaced within a specified range determined by an urban/rural classification. The displacement is done randomly in terms of direction and distance up to 5 km for the rural clusters, and 2 km for urban clusters. A 10 km distance-up is applied for one percent of rural areas.[19]

To measure the real agricultural output (Y), I construct a Laspeyres index of production that aggregates the quantity produced of main cash and staple crops crops (cassava, maize, yam, beans, cocoyam, millet, oil palm, and rice) produced by household farms using proxies of prices in 2010 as weights. I also identify the other, minor crops grown under a category named “other crops”. I use unit values as proxies of prices. To calculate these proxies, I follow Aragón et al. (2019) and divide the value of sales by the quantity of each crop. I then calculate the median unit value of each crop at national level.

For the main agricultural inputs, I construct land input by adding up the size of plots harvested by households. Labor input is estimated by adding hired worked days to the number of days that all members of the household spend working on the household

[18] The post-planting and post-harvest visit calendars are shown in the Appendix.

[19] The reason for this modification of coordinates is to meet user interest in geo-referenced locations while preserving the confidentiality of sample households and communities.

farm. I use the endowment of each household as an instrumental variable, following the methodology of Benjamin (1992). Available land is the sum of the area of all plots to which a farmer has access, either by the distribution of the community or family, outright purchase, renting, or use free of charge. Labor endowment is the number of equivalent adults in households.

The survey also provides information on household characteristics and agricultural practices (age of head of household, literacy, an indicator of whether households own their land, use of fertilizers, herbicides, pesticides, and improved seeds). I use these as control variables in my main specification and robustness analysis.[20] In the robustness analysis, I also supplement household and agricultural practices data with a set of geospatial variables that help to control for other characteristics that could also affect the total agricultural productivity channel. These include long distances to main points (federal roads, main towns, main markets, state capitals, and border posts), mean rainfall levels and temperatures, soil characteristics (landscape type, level of toxicity, excess salt, workability, nutrient retention and availability, and oxygen availability to roots).

Table B1 in the Appendix presents summary statistics of the agricultural characteristics, households variables, and weather and terrain conditions. There are several relevant observations for my analysis. First, farmers have small scale operations with no substantial differences between the plot areas harvested and their total plots (the average total land harvested is 3.77 hectares and the average total plot size is 4.05 hectares, giving a figure of around 93%). Second, farmers use practices that can be described as subsistence farming, e.g. limited use of pesticides and herbicides. Table B2 presents summary statistics from the dataset which are not only restricted to agricultural workers (rural or urban, population, sex, age). Some of these variables may explain a drop in agricultural total productivity not associated with oil-spill pollution (dummy variables if an individual, male or female, is employed, semi-employed or hired in domestic production, works in agriculture, migration, literacy, secondary education, and own business). I also present some household variables that help me to explain a drop in labor productivity (dummy variable if any individual reports being ill in the last four weeks, number of days for which an individual reports ceasing to engage in any usual activity, number of total hours worked, and real employment income).

[20] https://microdata.worldbank.org/index.php/catalog/1002

The oil spills dataset

The second database used is The Nigerian Oil Spill Monitor, which provides geo-referenced data from January 2006 to December 2018 on oil spills registered by the National Oil Spill Detection and Response Agency (NOSDRA), the Nigerian environmental regulator. The Nigerian Oil Spills Monitor visualizes oil spills on an online map and allows data to be downloaded in a table.[21] The data prior to 2013 is not entirely well-referenced. In most cases, only the site location is provided. In these cases, I use the geocoding[22] tool from the geographical information system (QGIS)[23] to obtain their geographical coordinates. NOSDRA calls on the public to report oil spills by email or via a hotline, but relies on voluntary engagement and on the support of oil companies to provide data. The dataset reports some supplementary information, such as the estimated quantity of oil spilt, the cause, the area covered, and the quantity recovered among other items. However, not all oil spills are supported by this information. Oil companies may be willing to provide information if oil spills are caused by sabotage or theft, or through their own fault (as in the cases of pipeline corrosion, maintenance, human operational errors, and equipment failures). There are 11,981 oil spills recorded for the period analyzed, around 68% of them attributed to sabotage. Most of these oil spills are concentrated on pipelines close to onshore oil and gas fields in the South-south zone of Nigeria

Figure 1 illustrates the total number of oil spills per annum over my sample period. There are no suitable references for oil spills before 2006. The figure shows a steady increase in oil spills in 2013 but a sharp decrease in 2015. This last evidence is consistent with the drop in oil prices in Nigeria[24] from that year onwards, suggesting that one reason for this decrease in oil spills could be a decrease in the sabotaging of pipelines to steal oil.

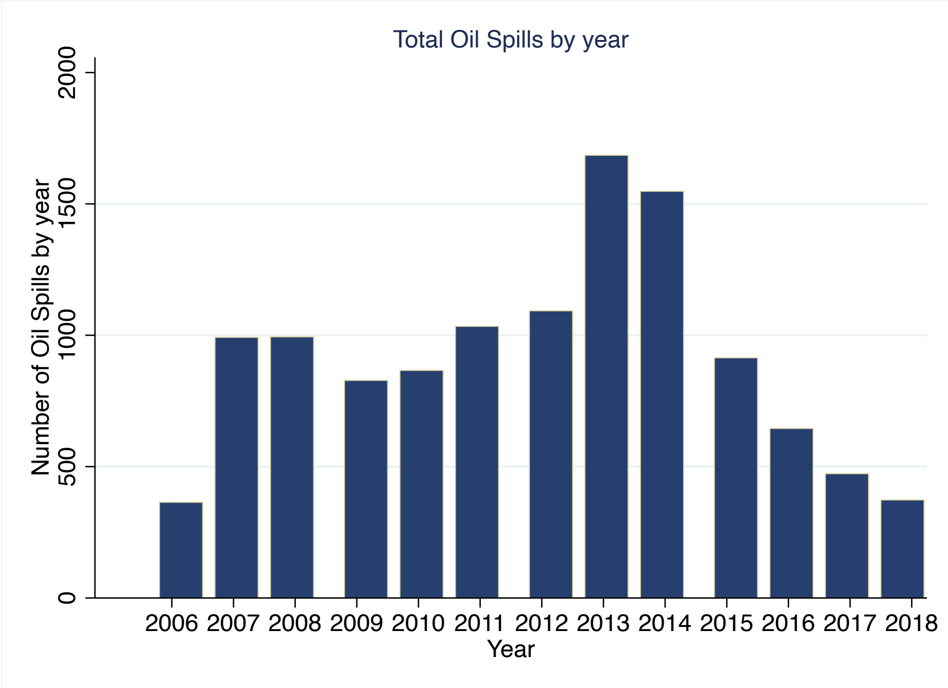
[21] The dataset of oil spills used in this paper was downloaded in January 2019.

[22] Specifically, I use the Geocode tool to geo-reference the exact site.

[23] QGIS is a user-friendly Open Source Geographic Information System (GIS) licensed under the GNU General Public License. QGIS is an official project of the Open Source Geospatial Foundation (OSGeo). It runs on Linux, Unix, Mac OSX, Windows and Android and supports numerous vector, raster, and database formats and functionalities. https://www.qgis.org/es/site/about/index.html

[24] See https://www.cbn.gov.ng/rates/crudeoil.asp, the Central Bank of Nigeria for historical oil prices in Nigeria.

Figure 1: Number of oil spills per year. Years: January 2006- December 2018.



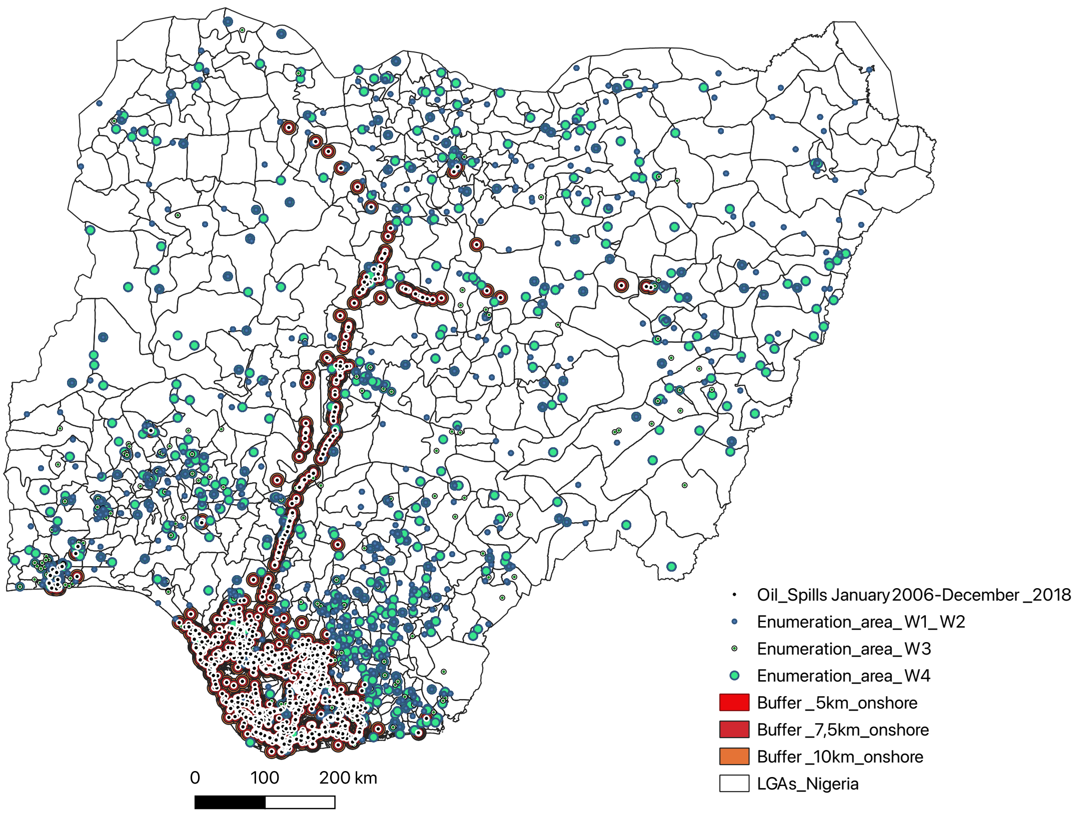
Sources: Own work based on Nigerian Oil Spill Monitor *NOSDRA*.data

To link oil spills and household data, I use the QGIS program mentioned above. Thus, I obtain the geographical coordinates for both oil spills and enumeration areas on a map. On average, each enumeration area contains ten households. I focus on the enumeration areas located near oil spills. Thus, I create buffers around oil spills that define proximity to them.[25] [26] Figure 2 is a map of Nigeria showing the location of oil spills and the enumeration areas for each wave. The shaded areas are the union of all buffers within radii of 5, 7.5, and 10 kilometers of all oil spills in my dataset.

[25] I present results up to 5 km in the main text and 7.5 km in the Appendix. Recall that the coordinate modification strategy in the Nigeria GHS-panel surveys means that up to 99% of households are located within a buffer zone of 5 km from the reported coordinates.

[26] Bruederle and Holder (2019) focus on mothers who live in clusters provided by the DHS survey at a reported distance of less than 10 km from the closest oil spill. According to Aragóon and Rud (2016), p. 1982 "[..]..using satellite imagery it is found that the concentration of (NO2), an indicator of air pollutant, is higher in locations near mines and declines with distance". They define buffer zones of 20 km from mine sites as the mining area. Nevertheless, oil spill pollution differs from that caused by mines, and Nigerian oil is very light in chemical composition with high levels of evaporation. Thus, I begin my analysis with the nearby locations. In a robustness analysis, I also consider longer distances (20km to 50 km).

Figure 2: Location of total onshore oil spills, enumeration areas, and buffers at 5km, 7.5km and 10km.



Sources: Own work based on Nigerian GHS-PANEL data and Nigerian Oil Spill Monitor *NOSDRA*.data

Conflict Data

I also use spatially explicit data from The Armed Conflict Location and Event Dataset (ACLED Dataset) (Raleigh et al., 2017). These data cover different countries and periods. Events are collected from various sources, including humanitarian agencies, research publications, and local, regional, and international press. In each dataset, the unit of observation is the event. They contain latitude and longitude coordinates and the exact day (in most cases) of conflict events. I construct a dummy variable that indicates whether any conflict event causing at least one fatality took place within 25 kilometers[27] of the reported enumeration area during the sample period of the Nigeria GHS-Panel, i.e. from January 2009 to December 2018.

[27] I follow the approach of Bruederle and Hodler (2019).

3.2 Empirical Strategy

The objective of my empirical analysis is to determine the extent of the effects of oil-spill pollution on agricultural activity. To that end, I estimate a production function and assess the impact of oil spills that occurred in zones near farms on total factor productivity A.

I follow the empirical implementation method set out in Aragón and Rud (2016), who study the expansion of mining activities in Ghana, for two reasons: First, pollutants affect agricultural total factor productivity similarly in both cases; and second, their impact may be higher on the areas near the main sources of pollution. However, in Aragón and Rud (2016), the pollution comes from mines, which are in fixed locations and pollute the air continuously. In this paper, oil spills are rare events on pipelines that might recur over time,[28] and their effects are also persistent in the environment.

I also assume the following agricultural production function:[29]

$$Y_{i,e,t} = A_{i,e,t} X_{i,t}^{\alpha} L_{i,t}^{\beta} e^{\epsilon_{it}} \tag{2}$$

where Y is the agricultural output of farmer i, in enumeration area e, in time period t. $A_{i,e,t}$ is total factor productivity, $X_{i,t}$ and $L_{i,t}$ are the actual land and labor inputs. $e^{\epsilon_{it}}$ captures unanticipated shocks not related to oil spills, which are by definition uncorrelated with input decisions by farmers. Finally, α and β are the input shares of land and labor respectively. Farmers may need other inputs for production, such as fertilizers, herbicides, animals, etc., but these are not commonly used so, following Aragón and Rud (2016), I decided to exclude them from the benchmark model, though I do include them later in my robustness analysis.

Total factor productivity $A_{i,e,t}$ is composed of three factors: O_{et} a function of the total number of oil spills in the proximity of enumeration area e before time t; the heterogeneity of farmers ($\chi_{i,t}$) and time-invariant environmental conditions and the local economy (ν_e). Hence, $A_{i,e,t}$=exp($\delta O_{et} + \chi_i + \nu_e$). The parameter of interest is δ. If $\delta < 0$, oil spills are

[28] An examination of the oil spills in the dataset shows several spills in nearby geographic coordinates and in the same year, which hence affect the same farms. Thus, my oil spill pollution variable reflects the number of oil spills near a given location and year.

[29] I am assuming Cobb-Douglas technology for the sake of simplicity and because I follow the methodology of Aragón and Rud, 2016; and Restuccia et al. 2006.

affecting total factor productivity, but if $\delta = 0$, the effect of oil spills could be transmitted via the input competition channel through prices or availability of inputs.

A limitation arises when I only approximate accumulated pollution via the number of oil spills. The amount of oil spilled and the spillage rate are key determinants of the severity of the consequences (Chang et al. 2014). Madu et al. (2017) find no correlation between these variables, but the estimated volume of the oil spills variable reported by NOSDRA[30] is incomplete for all events and implies an observation loss of approximately one-third. Thus, I decided not to consider this variable in my main model, although the volume spilled could be key in quantifying the damage to the environment.[31] However, a histogram of the variable shows that the density of its probability is concentrated in a small quantity of oil spills: There are few medium or large oil spills.[32]

There are also some empirical challenges: Places around the pipelines, and hence locations near oil spills could have permanent differences in productivity. In particular, the sharp increase in oil spill data in 2013 could indicate that data from previous oil spills was not fully reported.[33] As Aragón and Rud (2016) point out, when I estimate the coefficient of interest, this omitted variables problem may lead to endogeneity issues.[34] To avoid such issues, I unify the buffer zones around oil spills as defined above to create an oil spill area.[35] I also use time variation in the repeated cross-section to compare differences in productivity in oil-spill and non-oil-spill areas. As highlighted by the aforesaid authors, this is basically a difference-in-difference methodology with continuous treatment. In this case, proximity to oil spills defines the control group and the intensity of treatment is the proxy for the estimated quantity of oil lost. Taking this approximation, I assume that the trend in output in both areas would have been similar without the presence of oil spills. In fact, most pipelines and oil spills are concentrated in the Niger Delta, a specific area where oil exploration has impacted the entire ecosystem.[36] Assuming this empirical strategy, the

[30] NOSDRA gives the amount in barrels reported as spilled by each company. However, the time-series data of both the estimated quantity of oil lost and the estimated quantity of oil recovered in each location are incomplete.

[31] I present additional results with this variable in my robustness analysis.

[32] See Figure A3 in the Appendix.

[33] See Bruederle and Hodler (2019).

[34] Zabbey et al. (2007) pg. 3, recall that "[..] An estimated 10 m to 13 m tons of hydrocarbons have been reportedly spilled into the Niger Delta over the last 50 year. During this period over 77% of spilled hydrocarbons are not recovery". Thus, I can assume that the number of oil spills is greater than reported.

[35] See Figures A4 and A5 in the Appendix.

[36] Many activities associated with oil extraction have affected the Delta area negatively. In particular, building infrastructures for oil exploration (such as access roads and canals, resulting in deforestation), laying pipes and gas flaring seriously damage the environment. For instance, gas flaring introduces toxic pollutants (such as PAHs and toxic metals, especially vanadium) into the air. In the Niger Delta, gas flaring facilities

variable O_{et} takes the value of 0 in the enumeration areas farthest from the defined oil spill area.

Second, in estimating my production function, both agricultural output and input choice could be affected by productivity being simultaneously determined. In this case, unobserved heterogeneity in productivity is reflected in the error term, creating an endogeneity problem in the estimation of input coefficients.

I address these issues in several ways. First at all, I use variables such as observable characteristics of farmers as proxies for heterogeneity, χ_i and I take LGAs fixed effects to capture differences in average output due to heterogeneity in the local economy. Taking logs in the agricultural production function, I estimate the following equation:

$$y_{i,e,d,t} = \alpha x_{i,t} + \beta l_{i,t} + \delta O_{et} + \nu Z_i + \kappa Oilspillarea_e + \mu_d + \theta_t + \varepsilon_{i,e,t} \tag{3}$$

where $y_{i,e,d,t}$ represents the log of the agricultural output of farmer i, in enumeration area (village or urban neighborhood) e, in LGAs d, in period t, $\mathrm{x}_{i,t}$ and $l_{i,t}$, are the log of land and labor for household i in period time t, respectively. The variable Z_i is a set of farmer's control set, O_{et} is the proxy for cumulative oil pollution in the proximity of the enumeration area; $Oilspillarea_e$ is a dummy that determines whether the land is within 5 kilometers of an oil spill, μ_d is a set of LGAs fixed effects, and θ_t is a set of time fixed effects. Finally, $\varepsilon_{i,e,t}$ is the corresponding disturbance term.

The above identification assumption means exploring the presence of some constrained farmers too. I estimate a standard IV model using input endowments as instruments for my observed use of inputs. As mentioned above, traditional farming is the main source of livelihood in Nigeria, which means that most farmers could have constrained access to the inputs market. As Benjamin (1992) points out, the greater the proportion of constrained farmers, the closer the correlation between household endowments and input uses. However, I can only use this approach in cases where correlation is strong enough and endowment affect output only through input uses and not through the productivity parameter A. That means that endowments are not conditionally correlated with the

are often close to local communities with no protection, leading to a high level of exposure to pollutants among households (Nriagu, 2011).

residual unobserved heterogeneity term, $\varepsilon_{i,e,t}$, which corresponds to the error term, $e_{i,t}$, heterogeneity in locations, and unaccounted farmers. Under this assumption, I can estimate my model with OLS regression. [37]

Finally, changes in agricultural productivity could be driven by other events that correlate in time and space with oil spills. Oil spills are likely to occur simultaneously with other events specific to oil production. Thus, following Bruederle and Hodler (2019), I use total factor productivity to compare the effect on agricultural output in the oil-producing states in the Niger Delta with the effect for agricultural output elsewhere. Oil spills are often also the result of vandalism. Sometimes military groups attack pipelines, which entails violence against civilians (U.S. Energy Information Administration, 2016). I use total factor productivity to compare the effect of oil spills on agricultural output for household farms close to conflict areas and households far from conflict areas.

Oil spills location and the oil spills pollution function

Oil-spill pollution is measured based on the location of oil spills near to households and, hence, near cultivated plots. As a proxy of oil-spill pollution, I use a function of the number of oil spills until period t. It is plausible that the enumeration areas near oil spills may receive the greatest impact both directly or indirectly through air and water contamination, because of the proximity of wetlands. As mentioned above, I consider the survey enumeration areas at a reported distance of less than 10 kilometers from the closest oil spill as my treatment group.[38] However, the effect on total agricultural productivity could fade away over time. In particular, I define the total oil spill function as follows:

$$O_{e,t} = \sum_{n=0}^{a_t} g(n) * Total\_oilspill_{e,t-n} \tag{4}$$

[37] In a robustness analysis in the Appendix, following Aragón and Rud (2016), I also consider the possibility that endowments may be correlated with the error term, $\varepsilon_{i,e,t}$. In that case, the exogeneity assumption in the IV strategy does not entirely apply. That situation emerges for more productive farmers who have systematically larger plots or households. To solve this issue, I apply the partial identification strategy used by Nevo and Rosen (2012), which implements an imperfect instrumental variables (IIV) strategy to identify a set of parameter values instead of point values. Recall that IV strategy relies on identifying point values. IV permits partial correlation with the error term. In particular, this approach implies that "[..] (i) the correlation between the instrument and the error term has the same sign as the correlation between the endogenous regressor and the error term, and (ii) that the instrument is less correlated with the error term than is the endogenous regressor" (Nevo and Rosen, 2012, p. 659). Given that I use the same instrumental variables as previous literature, I carry out the same exercise to check the validity of my instrument variables.

[38] The results show that the best specification in this case is to choose buffer zones of up to 5km. Taking enumeration areas up to 10 kilometers from oil spills is not significant.

where $Total\_oilspill_{e,t-n}$ is the number of oil spills close to an enumeration area e, t is the period of the wave, n is the number of years before each wave, and a_t is the total number of years before each wave. [39] The persistent effect is defined by $g(n)$. This function takes different specifications depending on how persistent is the effect of oil spills on soil quality and human capital is. In particular, in my benchmark model, $g(n) = exp^{-n}$.

With this formulation, I consider the possibility that oil spill incidents during the period of the wave have the most impact on total factor productivity. That means that contemporaneous oil spills impact agricultural output strongly in present crop seasons. I also consider that productivity depends on previous oil spills that may have impacted both the quality of soil and human capital. However, these impacts decline exponentially over n years. The main processes that influence the degradation of oil spills include evaporation, auto-oxidation, and microbial degradation. The first known model to describe a process of decomposition of organic matter is the simple exponential model, which was initially proposed by Jenny et al. (1949) and discussed in detail by Olson (1963). I decided to take this approximation to create my oil spills pollution variable, given that the oil would affect soil quality at a lower rate over time. In particular, the original function is: $X = X_0 e^{-kt}$, where X is the amount of litter remaining at the time t from an initial amount X_0. For the sake of simplicity, I consider the value of k to be 1. To the best of my knowledge, there are no previous studies that determine how many years this effect will persist in the soil.[40] It probably depends on the degree of evaporation, the chemical composition of the hydrocarbon contaminants, the physical characteristics of the terrain, weather issues, other environmental factors (including PH and soil aeration), and clean-up aspects. To the best of my knowledge, there is no way of determining how the process works or how long chemicals of these types can affect the quality of land. Thus, I consider all oil spills up to the last day of the harvest survey for each wave as a cumulative, persistent effect on total factor productivity. Nevertheless, I also use different measurements in the robustness analysis. For example, oil spills may have the same impact on soil independently of the year when they happen. Thus $g(n) = 1 \rightarrow \forall n$.[41] Another approach is to identify the number of oil spills that affect land from the beginning of the planting period to the end of the harvest period given by each survey.

[39] In particular, a_t could be $a_{2011} = 5$, $a_{2013} = 7$, $a_{2016} = 10$, $a_{2018} = 12$.

[40] The UNEP report (2011) concludes that contamination persists for many years.

[41] In my analysis, the biota of soil quality could be lost for many years because of pollution. Another reason may be that oil spills could affect labor productivity permanently through chronic diseases suffered by the labor force. Thus, I decided to do a robustness analysis with no degrading effect.

$$g(n)=\begin{cases} 1 & \text{si } n=0 \\ 0 & \text{si } n>0 \end{cases}$$

In this case, the effect is not persistent over time. Finally, I consider oil spills that occurred up to five years before the planting period. In this last case, I apply the same number of years to each wave. Thus, $g(n)=\begin{cases} exp^{-n} & \text{si } n=0, 1,.....,5 \\ 0 & \text{si } n>5 \end{cases}$

In short, my cumulative oil spills function depends on the total number of oil spills near location e and on the year of each spill. Effects are always greater if oil spills occur during the year of each survey, given that they could affect both agricultural productivity and input uses in that year directly or indirectly. However, if there are events before each survey, agricultural productivity may probably be affected by their persistent effects on the quality of soil and human capital. The extent of that persistence over time will depend on how I parameterize the function g(n).[42][43]

Table 1 presents a simplified difference-in-difference estimation of the main variables, comparing mean values in all waves for farmers located in oil-spill and non-oil spill areas. The first observation is that in both areas the log of agricultural output decreases in 2012-2013 and in 2018-2019. However, the impact is stronger for oil-spill areas in 2012-2013. In that period the number of oil spills increased. In fact, there is a stronger significant difference in this variable when the two zones are compared. There is also a clearly significant difference in the use of labor input. Land harvested is slightly significant at 10%, but labor is negative and significant at 1%, suggesting an adjustment of this input in the spill area. Concerning household characteristics, I find clear evidence that less land is owned by farmers who live near oil spills. The head of the family also tends to be younger, but careful examination of this variable reveals that the significant impact may be due to the inclusion of new households in wave 4. However, these differences disappear in both cases when I control for other household characteristics. Finally, the greater use of fertilizer may suggest that farmers take action because of the pollution perceived on their land.

[42] See Table B3 in the Appendix for summary statistics of the oil-spill pollution variables.

[43] See Table B4 in the Appendix for data on collection dates of surveys and oil-spill incidents considered for each period for the main analysis and the second approach.

Table 1: Mean of main variables, by wave and location

VARIABLE	Within 5 km of oil spill				More than 5 km from oil spill				Diff. columns
	2010-2011 (1)	2012-2013 (2)	2015-2016 (3)	2017-2018 (4)	2010-2011 (5)	2012-2013 (6)	2015-2016 (7)	2018-2019 (8)	(4-3-2-1)-(8-7-6-5) (9)
Cumulative Oil Spill	1.065	2.619	1.128	7.042	—	—	—	—	—
Ln Real Agricultural Output	10.193	9.529	10.554	7.428	10.370	10.158	10.577	6.896	-0.02*** (0.005)
Land harvested (hectares)	0.783	0.920	1.177	1.027	7.233	3.596	3.121	1.311	0.08* (0.042)
labor (days)	154.870	78.335	299.809	708.565	215.904	218.284	381.436	725.299	-7.378*** (2.542)
N^o members in household	2.758	3.823	5.175	3.416	2.929	3.497	4.460	3.527	-0.023 (0.024)
Owner-occupied farmland (%)	82.185	80.539	66.212	72.868	81.483	75.596	75.820	75.906	-0.01*** (0.003)
Age of head of family (years)	48.384	60.368	56.135	45.632	50.649	53.123	54.416	49.604	-0.18** (0.071)
Literacy (%)	75.361	50.913	58.035	80.670	52.353	58.207	55.509	69.329	0.000 (0.001)
Fertilizers	0.429	0.155	0.218	0.400	0.442	0.459	0.599	0.483	0.00** (0.001)
Improved_Seed	1	0.952	0.989	1	0.977	0.958	0.977	1	-0.00 (0.000)
Small Business	0.699	0.346	0.428	0.472	0.524	0.444	0.469	0.539	-0.01 (0.005)
Observation	38	54	156	188	1,639	1,235	1,757	1,881	

Notes: Columns 1-8 report mean values for the sub-samples of farmers less and more than 5 kilometers from an oil spill for each wave of the Nigerian GHS. Means are estimated using simple weights. Column 9 displays the coefficient of the regression estimate for each variable. This is obtained by regressing each variable on Cumulative Oil Spill and a dummy for being 5 kilometers from an oil spill. As in the baseline regressions, standard errors are clustered at LGA level. Fixed effects are included, but no control variables. By definition, Cumulative Oil Spills more than 5km from an oil spill are zero in all periods. The total number of observations is 5,998. *** p<0.01, ** p<0.05, * p<0.1. The standard errors are in parentheses.

4 Main Results

4.1 Effect on Agricultural Productivity

This section provides the results for the main hypothesis, evidencing that oil spills near farmers are associated with significant reductions in agricultural productivity. I start with the baseline specification using the total cumulative oil-spill function in locations up to 5 kilometers from the oil spill.

Table 2 presents the main results. Column 1 examines the link between agricultural output and the proxy for total accumulated oil-spill pollution in nearby locations, without controlling for input use. I find that link to be negative and significant, and consistent with oil spills affecting agricultural output through both pollution and input competition as the model developed by Aragón and Rud (2016) suggests. Next, I explore the channels likely to be driving the link. Column 2 estimates the agricultural production function defined in (1) with the OLS model, while column 3 estimates the 2SLS regression using input endowments. [44] All regressions include a set of controls for farmers, wave dummies, and

[44] The results of the First Stage show a positive, significant correlation between inputs and input

Table 2: Oil spill pollution and Agricultural Productivity.

VARIABLES	Ln Agricultural Output			LnYield	LnYield_Cas	LnYield_Maize
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative_OilSpill	-0.0236*	-0.0234***	-0.0273***	-0.0288***	-0.0373*	-0.0152
	(0.013)	(0.006)	(0.007)	(0.010)	(0.022)	(0.019)
LnLand		0.2572***	0.5384***			
		(0.024)	(0.171)			
Lnlabor_days		0.2001***	0.2954*			
		(0.025)	(0.153)			
Estimation	OLS	OLS	2SLS	OLS	OLS	OLS
Observations	6390	6130	6114	9051	1177	2564
R-squared	0.612	0.640	0.618	0.308	0.300	0.294
Waves dummies	YES	YES	YES	YES	YES	YES
LGAs fixed effects	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being 5 kilometers from an oil spill. Controls on farmers: columns 2 to 4 and column 6 give household head age and literacy and an indicator of whether farms are owner-occupied. Denotes significance at *** p<0.01, ** p<0.05, * p<0.1 Column 3 is estimated using 2SLS. Column 3 is estimated using 2SLS. The instruments excluded are the log of the area of land managed and the log of the number of equivalent adults in the household.

LGA fixed effects. All specifications use cluster errors at LGA level and sample weights to account for both autocorrelation spatial patterns and sample design.

Results suggest a negative link between the presence of nearby oil spills and agricultural output once input use is controlled for.[45] Following the identification strategy, I interpret these results as evidence of a reduction in agricultural productivity. Thus, oil-spill pollution affects the agricultural sector negatively in the areas affected. To further quantify the results in column 3, following the standard procedure in the literature, I find that an increase of one unit of the accumulated total number of oil spills leads to a decrease in agricultural production of around 2.73%; alternatively, an increase of one standard deviation in the measuring of accumulated oil spills is associated with a reduction of almost 8% in agricultural productivity. [46]

Columns 4 to 6 show the effect of oil spills on crop yield, which is defined as physical output per unit of land. Column 4 shows the sum of the yields from cassava, maize, and yam; column 5 those from cassava; and column 6 those from maize. These are the main crops in both oil-spill and non-oil spill areas. Crop yield is a standard measurement of agricultural productivity that abstracts from deflation and output aggregation issues. However, it gives no information about whether changes in agricultural productivity are

endowments. See Table B5 in the Appendix. As a further check, in Figure A7 and Table B6 in the Appendix I present the estimations using the imperfect instrumental variable (IIV) approach. Figure A7 shows that the effect on residual productivity is negative in more than 96% of all combinations.

[45] Using the 2SLS estimation in column 3, the results of α and β do not reject the null hypothesis of constant returns to scale at 5%.

[46] The impact is computed as marginal effects as follows. It is given by the standard deviation times the estimated coefficient multiplied by 100.

generated by changes in inputs or in total factor productivity, A. In all cases, I estimate an OLS regression that includes controls for farmers and fixed effects. As expected, the results are negative, suggesting again that the effect of oil spills on agricultural productivity is negative and significant.

Spatial disaggregation

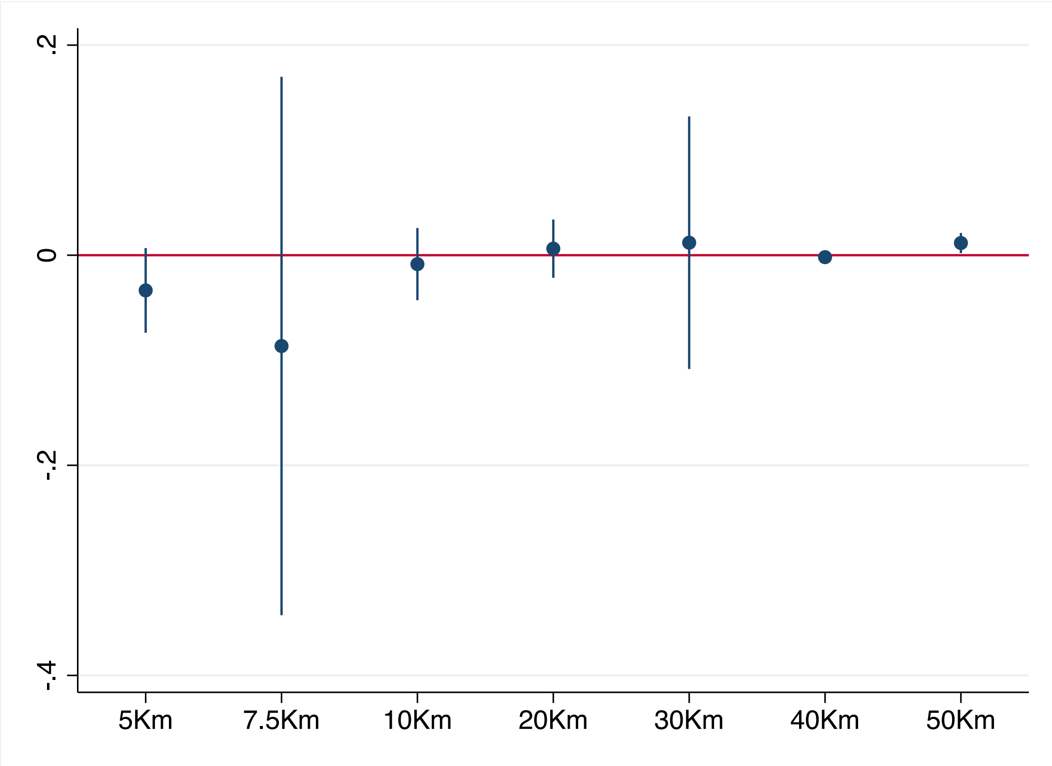
Recall that I consider areas within 5 kilometers of oil spills as being hit harder.[47] I now disaggregate the effects by distances between oil spills and farming plots. Specifically, I focus on how spatial proximity to oil spills affects the extent of their effect. To that end, I construct indicator variables for the geographical distance between the reported enumeration area and the closest oil spill. Thus, I replace $O_{e,t}$ by a linear spline of the main variable included in each distance bracket b.

In particular, it is replaced by: $\sum_b \gamma^{p,k} \sum_{n=0}^{a_t} exp^{-n} * Total\_oilspill_{e,t-n}^{p,k}$. This refers to the sum total of oil spills close to enumeration area e, within the distance brackets with lower and upper limit p, and k, respectively,[48] allowing for previous total oil spills in exponential decay. The estimates of $\gamma^{p,k}$ are presented in Table B7 and Figure 3.

[47] Remember that the effect of pollution is located in places near oil spills, although it could spread through water and filtration into the ground through rivers, and by air through the fires caused by oil spills.

[48] The distances are the following: 0-5km, 5-7.5km, 7.5-10km, 10-20km, 20-30km, 30-40km, 40-50km.

Figure 3: The Effect of Oil-Spill Pollution on Agricultural Productivity measured in Distances from Oil Spills.



Notes: This figure shows the estimates of $\gamma^{p,k}$ for the following values of p and k: 0-5km, 5-7.5km, 7.5-10km, 10-20km, 20-30km, 30-40km, 40-50km. Circles represent point estimates, while lines indicate the 95 percent confidence interval.

The effect of oil spills on total agricultural productivity is greatest for oil spills that occurred less than 7.5 km from the cluster location. The loss of productivity becomes smaller and positive in locations more than 10km from an oil spill. However, the confidence intervals are large, given that the number of enumeration areas considered is much lower for each individual treatment than for the combined treatment. Because of the larger decrease in total agricultural productivity in the 5km to 7.5 km interval,[49] I also include enumeration areas up to 7.5 km away as a focus for my treatment. I present the results in Table B8 in the Appendix. The table is organized in the same way as Table 2. The results show that the effect of cumulative, persistent oil pollution on locations within 7.5 km are still significant in most cases. However, the effect is not so strong as the previous choice. This could also be due to the random displacement of 1% of clusters, so I take this

[49]Columns 1 to 4 of Table B7 show a large decrease in total agricultural productivity in areas between 5 and 7.5 km away, but that difference disappears once additional variables are introduced into the model.

as a validation of the idea that I should focus on locations within 5 km of oil spills.[50] I repeat the exercise in Table B9 with locations 10, 20, 30, 40, and 50 km from oil spills. Surprisingly, I find that the effect is again negative and significant only at locations up to 30 km from oil spills. Thus, these results should be interpreted with caution because most oil spills occur closer to wetland areas, and the oil spilled could flush over the surface of the water and affect large areas. The design of the analysis with distanced buffer zones around oil spills may not be the best choice, and other types of analysis such as the closest household to oil spills along river courses could be studied.

4.2 Robustness

4.2.1 Additional Control Variables

Table 3 presents several checks on the robustness of the main model. First, following Benjamin (1992), I introduce variables to control for additional heterogeneity that could bias my results. First, I estimate OLS without land and labor variables but including controls for whether a farm uses fertilizers, pesticides, herbicides, and improved seeds. The prices of these inputs have a significant effect on labor demand, suggesting that they could be a substitute for this input (Benjamin, 1992). In Column 2 I also reintroduce the main input variables. Column 3 expands these specifications by adding an array of heterogenous trends to the enumeration area level. Specifically, I add indicators of distance trends (nearest federal road, nearest major market, border post on the main road, major towns, distance to the capital of the state of residence). In column 4 I also introduce variables that affect total productivity and could capture other confounding factors for productivity and quality of plots. In particular, I use mean temperature and rainfall, rooting, slope, nutrient retention, excess salt, oxygen supply to roots, toxicity, and workability. These last characteristics are important to control for the quality of plots independently of oil-spill shocks. All my results show that the negative effect of cumulative oil-spill pollution plus real agricultural production is still significant when possible confounding factors are included.[51]

[50] I do likewise with locations within 10 km. The results are not significant in any specification. These additional tables are available on request.

[51] In the Appendix, I present the results using a 2SLS estimator. The results are very similar to the OLS estimation. See Table B10.

Table 3: Additional checks

VARIABLES	Ln Agricultural Output				
	(1)	(2)	(3)	(4)	(5)
Cumulative_OilSpill	-0.0249*	-0.0238***	-0.0260***	-0.0224***	-0.0244***
	(0.013)	(0.006)	(0.006)	(0.007)	(0.007)
LnLand		0.2466***	0.2459***	0.2566***	0.2463***
		(0.024)	(0.024)	(0.024)	(0.024)
Lnlabor_days		0.1904***	0.1911***	0.1978***	0.1876***
		(0.025)	(0.025)	(0.024)	(0.024)
Fertilizers	0.3272***	0.2611***	0.2638***		0.2734***
	(0.079)	(0.076)	(0.075)		(0.074)
Pesticides	0.3404***	0.2196***	0.2156***		0.2102***
	(0.069)	(0.066)	(0.067)		(0.066)
Herbicides	0.1845*	0.0955	0.1067		0.1099
	(0.099)	(0.095)	(0.095)		(0.094)
Improved_Seeds	0.2547	0.1516	0.1612		0.1496
	(0.353)	(0.274)	(0.274)		(0.272)
Rooting				0.4526*	0.4450*
				(0.262)	(0.248)
Oxygen to roots				-0.0432	-0.0149
				(0.191)	(0.195)
Toxicity				0.5825	0.3611
				(0.548)	(0.558)
Excess salt				-0.5637	-0.4103
				(0.469)	(0.481)
Workability				-0.4470*	-0.4433*
				(0.246)	(0.237)
Nutrient_Retention				0.2565	0.2816
				(0.253)	(0.248)
Nutrient_Availability				-0.2581	-0.2556
				(0.222)	(0.218)
Mean temperature				0.0137	0.0173
				(0.015)	(0.015)
Mean rainfall				-0.0001	0.0001
				(0.000)	(0.000)
Slope				0.0123	0.0103
				(0.017)	(0.017)
Estimation	OLS	OLS	OLS	OLS	OLS
Observations	6324	6076	6076	6130	6076
R-squared	0.618	0.643	0.643	0.641	0.644
Waves dummies	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 5 kilometers of an oil spill. Controls for farmers are as follows: columns 1 and 2 show age of head of household and literacy (an indicator of whether a household owns its farm plot). Columns 3 to 5 show indicators from time trends with distances to federal road, main towns, main markets, states capitals, and border posts on the main road. Significance levels are denoted as follows *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

4.2.2 Alternative Measurements of Oil Pollution

As mentioned above, my oil pollution measurement is based on the hypothesis that the impact of pollution on agricultural productivity is greater in the present crop season. The variable is also cumulative and persistent, although it fades over time with exponential decay. I also use the total number of oil spills rather than the estimated quantity of oil spilt in terms of barrels lost to formulate my proxy of cumulative, persistent oil pollution. My findings, however, may be affected by these choices. Therefore, as an additional robustness check, I examine the extent to which the above results depend on the definition of oil spill pollution used to construct the main variable in the model.

First, I consider the three additional specifications for my function $g(n)$, in order to formulate $O_{e,t}$, present in the above section. Figure A8 and Table B11 in the Appendix present the results for each additional measure. As can be seen, all results are negative and significant in OLS estimates on the agricultural output. Results show that the consequences of oil pollution are greater during the crop seasons for each wave. However, the effect is also persistent over time.

As a second approach, I reformulate all the oil pollution proxies used[52] with the estimated quantity of oil split, measured in barrels.[53] Figure A9 and Table B12 present the estimation results with different measurements of oil pollution based on this variable. In qualitative terms, the results are quite similar to those presented in Figure A8 and Table B11. The effect of cumulative oil pollution on both agricultural production and crop yields is negative in all regressions. However, if I only consider the estimated number of barrels lost in the year of each wave, the variable is insignificant in all cases.

[52] Recall that the volume of oil lost is an important indicator of environmental damage, but one third of the data for the variable is missing, which could bias my results.

[53] Specifically, the cumulative oil pollution proxy is formulated as follows:

$$\mathrm{O}_{e,t} = \sum_{n=0}^{a_t} g(n) * Total\_estimated\_quantity\_barrels_{e,t-n} \quad (5)$$

where $Total\_estimated\_quantity\_barrels_{e,t-n}$ is the estimated quantity number of barrels spills close to enumeration area e, t is the year of the wave, n is the number of years before each wave, and a_t is total number of years after each wave. For this variable, $g(n) = exp^{-n}$. In addition, I construct the following oil pollution measurements: Volumen of OilSpill (barrels) measures the estimated number of barrels lost nearby a location in the year of each wave. Thus, $g(n) = \begin{cases} 1 & \text{si } n = 0 \\ 0 & \text{si } n > 0 \end{cases}$. FCum_Volumen of OilSpill (barrels) measures cumulative oil spills that persist for only five years, $g(n) = \begin{cases} exp^{-n} & \text{si } n = 0, 1, \ldots\ldots, 5 \\ 0 & \text{si } n > 5 \end{cases}$. TCum_Volumen of OilSpil (barrels) is the total estimated number of barrels spilt near locations up to the last day of the harvest survey for each wave, g(n)=1.

5 Possible Confounders and Alternative Explanations

I interpret the above findings as a credible channel through which oil-spill pollution has affected agricultural productivity. In this section, I explore two possible confounders and four plausible alternative explanations.

5.1 Possible Confounders

I test whether the loss of productivity could be caused by events in just one part of the country. The General Household Survey divides Nigeria into six geopolitical zones.[54] In all zones, surveys report enumeration areas affected by oil-spill events. In the Niger Delta area there could be events linked to oil operations and extractions that drive this loss of productivity but are not exclusively oil-spills per se. [55] Moreover, violent events could also lead to a loss of agricultural output, affecting both agricultural total factor productivity and labor use. Specifically, I take the approach in Bruederle and Hodler (2019) by considering cluster locations within 25 km of conflict events that involve at least one fatality during the period.[56]

First, I rerun my main regression six times dropping these geographical zones one by one. Columns 1 to 6 in Table B14 and Figure A11 present the coefficient estimates. My oil spill pollution measurement remains negative and statistically significant in five specifications, the exception being when I exclude the South-South zone, which covers most of the Niger Delta. Column 7 in Table B14 shows the results when I restrict the data to just the Niger Delta area. The effect of oil pollution there is greater than when I analyze oil spills throughout the country.

These results may indicate that productivity losses could be driven by events in one specific zone. However, this interpretation should be taken with caution. The percentage of farmers who have suffered an oil-spill event is around 25% in the Niger Delta, but just 4% elsewhere. To clarify these results, I also conduct an additional test to examine whether oil spills outside the Niger Delta area have not impacted agricultural output. I address this issue by relaxing the baseline specification and comparing the effects of oil spills on

[54] Recall that Figure A2 in the Appendix gives a map of Nigeria showing its boundaries and geopolitical zones.

[55] Figure A11 in the Appendix gives a map of the Niger Delta region.

[56] Figure A12 in the Appendix gives a map of conflicts and buffer zones within 25 km of conflicts.

agricultural productivity in and outside the Niger Delta region. [57] The results in column 8 of Table B14 and in Figure 4, confirm that oil spills pollution affects both the Niger Delta and other regions. However, the effect is greater in the Niger Delta area, where both the number and the persistence of oil spills are higher.[58]

Second, I explore whether my results could be driven by violent conflicts. 91 percent of the enumeration areas have suffered both oil spills and violent conflict. Columns 1 and 2 of Table B15 show the figures for oil-spill pollution with the database constrained to conflict areas and non-conflict areas, respectively.[59] In both estimates, the results are negative and significant, confirming that oil-spill pollution is a channel that reduces agricultural productivity. Column 3 of Table B15 estimates the main specification in the same way as column 8 of Table B14, to compare the effects of oil-spill pollution in conflict and non-conflict areas. The result, which is also presented in Figure 4, shows that the effect of cumulative oil-spill pollution is far greater for farms outside conflict areas. This difference is however not statistically significant at 5%.[60] I therefore conclude that loss of total agricultural productivity in farms close to oil spills is not only a result of violent conflict.

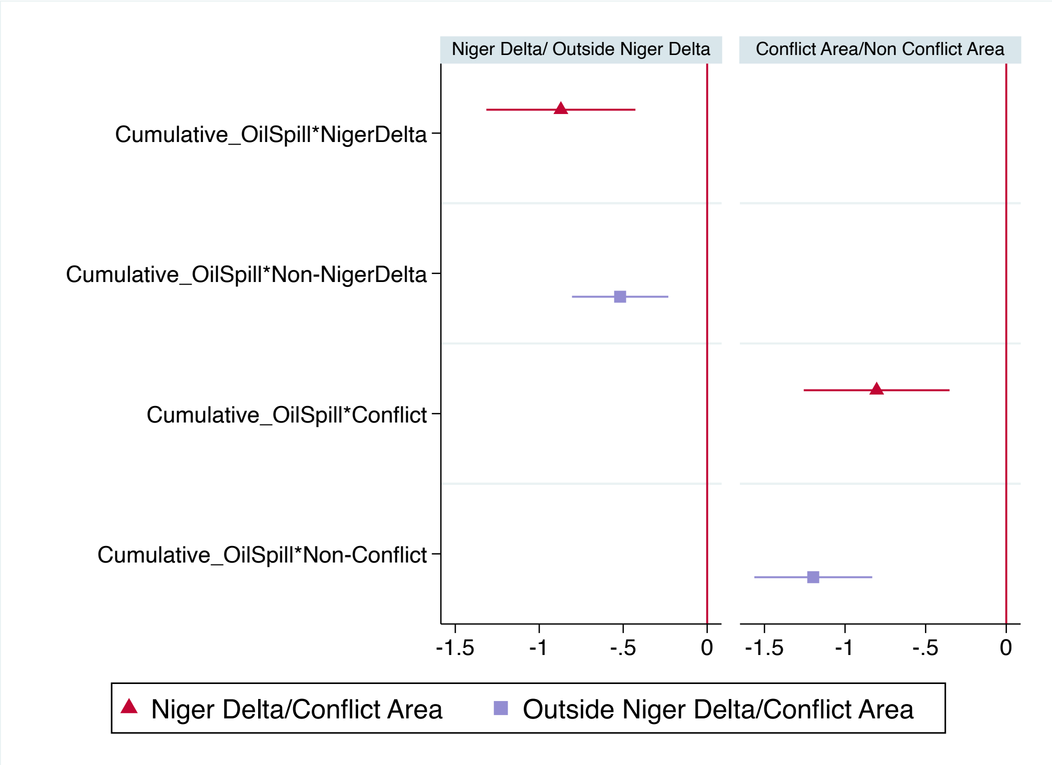
[57] To that end, I estimate my model with panel data from wave 1 to wave 3, which enables me to control for enumeration area time trends in the main specification. The inclusion of enumeration area time trends enables me to check that local economic, social or political developments are not driving the consequences of this difference between locations. Recall that I use a repeated cross-section in my analysis because longitudinal weight, which includes wave 4, is not currently available, which means that I cannot apply enumeration area time trends.

[58] Column 8 also displays the p-value of the test of equivalence.

[59] In both estimations, I decided to drop the North-East geographical zone because the Boko Haram crisis, which is located in this area, could bias the results. Boko Haram, led by Abubakar Shekau, is West Africa's most active and lethal actor. Since 2009, events involving it have numbered more than 2,350 and it has been linked to more than 27,000 fatalities. See also Figure A12. For more information, see the ACLED website: https://acleddata.com/crisis-profile/boko-haram-crisis/.

[60] Column 3 also displays the p-value of the test of equivalence.

Figure 4: Results for possible confounders: Niger Delta area and Violent Events.



Notes: The left panel shows the coefficient estimates from a linear regression of agricultural output on the interaction terms: $Cumulative\_OilSpill * NigerDelta\_area$, and $Cumulative\_OilSpill * Non - NigerDelta\_area$, controlling for LGAs, year time effects and ea-specific time trends. The right panel shows the coefficient estimates from a linear regression of the agricultural output on the interactions terms: $Cumulative\_OilSpill * Conflict\_area$, and $Cumulative\_OilSpill * Non - Conflict\_area$, controlling for LGAs and year fixed effects and ea-specific time trendss. The sample is the same as in the main specification (Table 2, column (4)). Geometric figures represent point estimates, while the horizontal lines represent 95% confidence intervals. Standard errors are adjusted for clustering at the LGAs level.

5.2 Alternative Explanations

My next step is to consider alternative explanations for the drop in total factor productivity following the approach in Aragón and Rud (2016). As mentioned above, oil and gas companies may demand local inputs (land or labor) for oil operations. For example, extractive companies could appropriate farmland to build additional infrastructures. In my estimations, I disregard all households that have been displaced. It is not possible to determine why these families decided to migrate, so farm reallocation could be a reason. Thus, this is not a plausible channel for explaining the drop in agricultural output.

Second, the fall in agricultural productivity might merely reflect changes in the composition of agricultural workers. For example, if the effect of oil pollution is permanent, members of the household could look for additional income by working in other sectors. Thus, Table 4 shows whether oil-spills are related to changes in various perceptible population characteristics. Columns 1 and 2 look at the probability of a working-age individual (male or female) being employed, semi-employed or hired in domestic production. Column 3 examines the probability of a worker being employed in agriculture. I would expect a negative correlation if there is an occupational shift towards non-agriculture activities. Columns 4 and 5 look at the demographics of agricultural workers and short-term mobility. Column 4 shows the probability of a worker being a prime-age male (20-40 years), while column 5 proxies the variable of migration with an indicator of whether any member of a household has been away for more than 30 days. Finally, columns 6 and 7 examine approximate measurements of the human capital of agricultural workers, such as literacy and having completed secondary school. These last measurements are informative because I am assuming that farming ability is positively correlated with education level. In Table 2 literacy is associated with an increase in total factor productivity and agricultural output. However, in Table 4, I find no significant evidence of any change in population characteristics except in the first and the fourth columns at the 10% level.

Table 4: Population Characteristics

VARIABLES	AnyWork (1)	AnyWork (2)	Works in agriculture (3)	Male_prime_age (4)	Migration (5)	Literacy (6)	Secondary (7)
Cumulative_OilSpill	-0.0058*	0.0056	-0.0036	0.0278*	-0.0184	-0.0220	-0.0138
	(0.003)	(0.005)	(0.003)	(0.015)	(0.018)	(0.016)	(0.028)
Sample	Males in working age	Female in working age	All workers	Agricultural workers	Agricultural workers	Agricultural workers	Agricultural workers
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Observations	16878	18204	35112	6976	11728	11776	7399
R-squared	0.428	0.323	0.293	0.300	0.093	0.245	0.152
Waves dummies	YES	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are shown in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 5 kilometers of an oil spills, an indicator or ecological zone and urban area. "Any Work" is a dummy variable that takes a value of one if an individual (male or female) is employed, semi-employed or hired in domestic production and 0 otherwise. Working age is between 15-65 years. "Works in agriculture" is a dummy with a value of one if an individual works in agriculture as a producer or laborer and 0 otherwise. "Male of prime age" is a dummy that takes a value of one if an individual is a male between 20 and 40 years old. "Migration" is a dummy variable with a value of one if any member of a family has been away for more than 30 days and 0 otherwise. "Literacy" is a dummy denoting whether an individual has literacy skills. "Secondary" is a dummy variable that denotes whether an individual has completed secondary schooling. Farmer controls include: age of head of household and literacy, and an indicator of whether the household owns its farm plot. Columns 1 to 3 include additional controls: age, age^2, literacy, and household size. Significance is denoted as follows: *** p<0.01, ** p<0.05, * p<0.1.

Another alternative reason that may explain a drop in agricultural productivity is related to weak property rights in Nigeria and the Land Use Act of 1978. The LUA replaced the previous plural land tenure system in Nigeria, with the idea of bringing consistency to the Nigerian land system. However, the risk of expropriation did not disappear under this legislation, and failure to provide compensation for oil pipeline failures is an important issue in this area. Thus, I first check whether there is a change in land ownership to examine property rights. Households could make less agricultural investments in rented farms. Moreover, farmers who have suffered oil spills before or who are located close to pipelines could use fertilizers or improved seed to minimize the effects of oil-spill pollution.

Finally, other channels could also be important in accounting for falls in agricultural output. For instance, a local oil operation boom may change the composition of workers, and the non-farming sector is also gaining significance in Nigeria (for example, in the South West region household enterprises account for half of all jobs (World Bank, 2015)). I am unable to examine the first channel directly due to lack of data, but I analyze the second by looking at whether any member of the family owns or manages a non-farm enterprise at least one year before the post-harvest visit for each wave.

Table 5: Agriculture Land Tenure and Practices. Small Business

VARIABLES	Owns_farm (1)	Fertilizers (2)	Improved_Seeds (3)	Own_Business (4)
Cumulative_OilSpill	-0.0141***	0.0025**	-0.0007*	-0.0034
	(0.002)	(0.001)	(0.000)	(0.005)
Observations	6704	6340	6328	6369
R-squared	0.304	0.569	0.393	0.266
Waves dummies	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES

Notes: Robust standard errors are shown in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being 5 kilometers from an oil spill. "Owns_farm" is a dummy that takes value of one if the household owns land and 0 otherwise. "Fertilizers" and "Improved_Seeds" are dummies that take a value of one if farmers use chemical fertilizers or improved seeds and 0 otherwise. "Own_ Business" is a dummy with a value of one if any member of the household owns or manages a non-farm enterprise and 0 otherwise. Farmer controls in columns 2, 3, and 4 include the age of the head of landowning households, literacy, and an indicator of whether a household owns its land. Significance is denoted as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5 shows the results. Firstly, I find that changes in land ownership are concentrated significantly in locations near oil spills. Thus, there could be a risk of expropriation in locations near pipelines, which could partly explain lower agricultural productivity. Concerning agricultural practices, I find a significant increase in the use of fertilizers that may suggest actions taken by farmers to offset the negative effects of oil pollution on land. However, this contrasts with my finding that the coefficient for improved seeds is again negative. Finally, I find no change in non-farm businesses, suggesting that there is no incentive to make additional efforts outside the farm in the treatment group.

The findings discussed above are far from conclusive, but they suggest that oil-spill pollution could be a plausible channel for explaining the decline in agricultural productivity in locations near oil spills. The effects of oil-spill pollution are found to be very local at first approximation, and the pollution caused by spills may explain this drop well. Authors such as Bruederle and Holder (2019) also remark on the significance of oil spills in the loss of health among children and adults. These findings are closely related to a drop in labor productivity..

6 Mechanisms

Section 2, above describes the mechanisms by which oil-spill pollution affects the total agricultural productivity factor. In particular, I consider three plausible channels: First, oil-spill pollution affects crop yields and health directly. Second, oil pollution deteriorates the quality of the soil and hence affects agricultural output. Third, oil spills affect human health, and hence labor productivity. In this section, following Aragón and Rud (2016), I discuss these arguments with the following augmented Cobb-Douglas production function:

$$Y = q_C(q_X X)^{\alpha}(q_L L)^{\beta} \tag{6}$$

where Y is agricultural output and X and L are the observable quantities of land and labor. q_X and q_L are input-specific quantity shifters, which capture factors such as quality of soil and labor productivity. q_C captures all other unobservable factors such as the crop health and yields. Thus, as analyzed above, oil-spill pollution might affect any of these factors.

In this framework, total factor productivity $A = q_C q_X^{\alpha} q_L^{\beta}$. That is the residual that I observe when I estimate agricultural output. My empirical analysis shows that the effect of oil spills reduces A but, as Aragón and Rud (2016) pointed out '[..]with the data at hand we cannot identify its effect on each component as this would require data on quality of soil, crop's health and labor productivity". Previous studies have demonstrated that oil pollution has a significant influence on soil properties and crop growth (Uquetan et al. 2017). Studies in the Niger Delta area find high prevalence rates for symptoms in human heath which are associated with oil spills in other parts of the world, including abnormalities in hematologic, hepatic, respiratory, renal, and neurologic functions (Nriagu, 2011). In any case, I can assume that not all the reduction in A is driven by a drop in labor productivity, q_L, but I cannot then identify the effects on soil quality or crop health, so I follow Aragón and Rud (2016); and conduct the same exercises that they do to assess the impact of pollution on labor productivity with additional tests.

First, I examine worker health indicators. I use self-reported data on the incidence of illness and cessation of usual activities from the Nigerian GHS-Panel data[61] to examine the link between these health measurements and my main oil pollution variable. I focus on working-age individuals (15-65) and splits between urban and rural populations.

Table 6: Oil Spill Pollution and Self-reported Illness

Variable	Ill in previous four weeks			Ln (Number of days off work)		
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative_OilSpill	-0.0062	-0.0092	-0.0006	-0.0017	-0.0034	-0.0026
	(0.005)	(0.006)	(0.004)	(0.003)	(0.003)	(0.007)
Sample	All	Urban	Rural	All	Urban	Rural
Observations	36675	9498	26196	23001	5916	16491
R-squared	0.070	0.055	0.080	0.225	0.178	0.252
Waves dummies	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are shown in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being 5 kilometers from an oil spill, and individual controls such as age, age^2, gender, an indicator of ecological zone and rural area. "Ill in previous four weeks" is an indicator that takes a value of 1 if an individual reports being ill during the last four weeks and 0 otherwise. This does not include injures. "Ln (Number of days off work)" is the log of number of days than an individual reports having ceased his/her usual activity. Significance is denoted as follows *** p<0.01, ** p<0.05, * p<0.1

Table 6 displays the results. In no case is there any evidence of an increase in the probability of being ill, or in the length of time for which activities are halted. I repeat the analysis with the total oil spills in the year of the wave, given that the survey reports short-term questions. Table B16 in the appendix shows the results. In this case, column 3 reveals a positive link between being ill and oil-spill pollution among workers in rural areas. Thus, a spill during the year may affect workers in rural areas, reducing their health

[61] Specifically, the questions on household surveys are the following:

1. During the last four weeks, have you suffered any illness or injury?

2. For how many days did you stop your usual activities?

In both cases, I center on illness, not injury.

and hence their labor productivity. This result is consistent with those of Bruederle and Hodler, 2019.

Second, I also examine the effect of oil-spill pollution on non-agricultural urban workers through the total number of hours worked and income from employment. If the effect of oil-spill pollution is transmitted through a reduction in labor productivity, drops in these variables could be observed. This group includes both employed and self-employed workers, and I assume the following: first, labor demand for urban workers depends on their productivity; and second, oil operations arising from oil spills neither increase labor demand in urban areas nor affect the urban labor supply. These last assumptions are plausible in the Nigerian employment marker. Given the capital-intensive nature of extractive sectors, their link with the rest of the economy is small, as is their contribution to job creation. Indirect jobs tend to be high-value-added jobs in the main urban areas, but this is probably not related to the issue at hand. (World Bank, 2015).

Table 7: Oil Spill Pollution and labor Outcomes for urban workers

Variable	Ln(Total hours worked)		Ln(Real Employment Income)	
	(1)	(2)	(3)	(4)
Cumulative_OilSpill	-0.0054	-0.0042	-0.0509***	-0.0552***
	(0.004)	(0.003)	(0.015)	(0.015)
Sample	All urban workers	Urban non-agric. workers	All urban workers	Urban non-agric. workers
Observations	4782	4369	2846	2788
R-squared	0.240	0.130	0.391	0.381
Waves dummies	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES

Notes: : Robust standard errors are shown in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being 5 kilometers from an oil spill and industrial dummies. Columns 1 and 2 include individual controls such as age,age^2, literacy, and household size. Columns 3 and 4 add additional controls in the form of the log of the total number of hours worked. All regressions exclude oil industry workers. Significance is denoted as follows *** p<0.01, ** p<0.05, * p<0.1

Table 7 presents the results. Although the no significant change is observed in the number of hours, there is a significant drop in income from employment that relates to a loss of labor productivity. In particular, an increase of one standard deviation in cumulative oil-spill pollution is associated with a reduction of around 11% in employment income in urban areas less than 5km from oil spills. I repeat the estimations in Table B17 with the oil-spill pollution variable during each wave (Oil spill). In this case there is clear evidence in both the total number of hours and employment income that confirms the link between the presence of oil spills near the locations and a drop in labor productivity.

7 Conclusions

This paper examines how oil-spill pollution affects total productivity and agricultural output. This is a type of externality that polluting industries impose on agricultural zones. Using geo-referenced data from the Nigerian Oil Spill Monitor and The Nigerian General Household Survey (GHS), I apply a difference-in-difference approach following the methodology in Aragón and Rud (2016). Previous literature has found that oil spills mean high levels of pollution in the affected regions, but that literature has not so far considered the cumulative, long-lasting effects of oil pollution on agricultural output through total factor productivity. I build up a novel variable that reflects this effect and obtain many interesting results. First, I find evidence that onshore oil spills near certain locations reduce their total factor productivity and agricultural output compared to other locations in the same LGAs but further from the spills. The effect for farmers is economically significant: there is a decline of approximately 2.73% in total factor productivity when an oil spill occurs nearby. Second, I show evidence that the effect does not disappear in the period studied. The empirical evidence suggests that the loss of productivity is also significant when only oil spills occurring in the year of the relevant wave are considered, but consequences are also persistent over time. Third, I find evidence that farmers are less likely to own their own land, and the drop in labor productivity leads to a decrease in both the health of workers in rural areas and employment income in urban areas. The number of oil spills recorded by the Nigerian Oil Monitor has fallen since 2014, but the results still indicate that it is necessary to pay attention to oil spills and their effects on nearby farmland. On the policy side, these results could be interpreted in terms of a need not only to prevent new oil spills but also to stress the effective clean-up of contaminated land, including surrounding environments. A system of compensation with community hosts for the economic losses suffered by farmers close to oil-spills is also required. The effects are long-lasting in time, so the system should offset these losses.

The main limitation of this paper is that I cannot exactly assess the relative importance of the various mechanisms through which oil-spill pollution may affect productivity, such as changes in soil quality. However, previous literature confirms effects on crop health.

Moreover, although I find evidence of a decrease in labor productivity, I cannot properly estimate the effect on the input competition channel or quantify the changes in input uses because of data limitations. These issues lie beyond the scope of this study, but examining them is a matter for further research. Distance is a logical indicator of the damage to soil quality caused by oil spills. However, most pipelines are very close to wetlands and oil pollution effects may spread to farms through rivers. In this case, an alternative empirical strategy based on the distance from oil spills to households along river courses may improve the study. Further research could consider such an analysis.

References

Akpokodje J. & Salau S. (2015). Oil pollution and agricultural productivity in the Niger Delta of Nigeria. *Environmental Economics* 6(4): 68-75.

Aragón F. M. & Rud J.P. (2016). Polluting Industries and Agricultural Productivity: Evidence from Mining in Ghana. *Economic Journal* 126(597): 1980-2011.

Aragón F. M. & Restuccia, D. & Rud J.P. (2019). Are small farms really more productive than large farms?*NBER Working Paper* No. w26331.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: testing for separation in agricultural household models, *Econometrica*, 60(2): 287-322.

Bruederle, A. & Hodler, R. (2019). Effect of oil spills on infant mortality in Nigeria. *Proceedings of the National Academy of Sciences* 116(12): 5467-5471.

Chang, E. S., & Stone, J. & Demes, K. & Piscitelli, M. (2014). Consequences of oil spills: a review and framework for informing planning. *Ecology and Society* 19(2): 2.

Currie, J., & Hanushek, E.A., & Kahn, E.M., & Neidell, M. & Rivkin, S.G. (2009). Does pollution increase school absences?.ÃÂ *Review of Economics and Statistics*, 91(4): 682-694.

FAO in Nigeria. http://www.fao.org/nigeria/fao-in-nigeria/nigeria-at-a-glance/en/ .

FAO (2010): "Addressing food insecurity in protracted crises" *The State of Food Insecurity in the World. Food and agriculture organization of the united nations.*

Frenske, J. & Zurimendi, I. (2017). Oil and ethnic inequality in Nigeria. *Journal of Economic Growth* 22: 397-420.

Graff Zivin, J.S. and Neidell, M.J. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7), 3652-3673.

Hanna, R. and Oliva, P. (2011). The effect of pollution on labor supply: evidence from a natural experiment in Mexico city, *NBER Working Paper.*

Inoni, O. & Omotor, D. G. & Adun, F.N (2006). The effect of oil spillage on crop yield and farm income in Delta State, Nigeria. *Journal of Central European Agriculture* 7(1): 41-48.

International Bank for Reconstruction and Development / The World Bank (2015): More, and More Productive, Jobs for Nigeria: A Profile of Work and Workers

Ite, A. E. & Harry, T. A. & Obadimu, C. O. & Asuaiko, E. R.& Inim I. J. (2018). Petroleum Hydrocarbons Contamination of Surface Water and Groundwater in the Niger Delta Region of Nigeria. *Journal of Environment Pollution and Human Health*, 6(2): 51-6.

Jayachandran, S. (2009). Air quality and early-life mortality: Evidence from Indonesia's wildfires. *Journal of Human Resources*, 44: 916-954.

Jenny, H. & Gesse, S.P.& Bincham, F.T. (1949) Comparative studies of the decomposition rates of organic matter in temperate and tropical regions. *Soil Science* 68: 419-432.

Maggs, R. & Wahid, A. & Shamsi, S.R.A. & Ashmore, M.R. (1995). Effects of ambient air pollution on wheat and rice yield in Pakistan. *Water, Air, Soil Pollution*, 85(3): 1311-1316.

Madu, C. N., & Kuei, C. & Ozumba, B.C. & Nnadi, V.E. (2018). Using the DPSIR framework and data analytics to analyze oil spillages in the Niger delta area. *Land Use Policy* 78: 78-90.

Marshall, F. & Ashmore, M. & Hinchcliffe, F. (1997). A hidden threat to food production: air pollution and agriculture in the developing world, *International Institute for Environment and Development (IIED)*, Working Paper.

Nwankwo, O. (2015) The Politics of Conflict over Oil in the Niger Delta Region of Nigeria: A Review of the Corporate Social Responsibility Strategies of the Oil Companies.

American Journal of Educational Research, 3(4): 383-392.

Nevo, A. & Rosen, A. (2012). Identification With Imperfect Instruments. *Review of Economics and Statistics* 94(3): 659-671.

NOSDRA. Nigerian Federal Government Agency. https://oilspillmonitor.ng

Nriagu, J. (2019). Oil Industry and the Health of Communities in the Niger Delta of Nigeria. *Encyclopedia of Environmental Health* 240-250.

Okonofua, B. (2011) Paths to Peacebuilding: Amnesty And The Niger Delta Violence. Georgia: Georgia State University.

Ojimba, T.P., (2011). Economic effect of crude oil spillages on crop farms in River State, Nigeria. *Global Journal of Pure and Applied Sciences* 17(2): 131-134.

Ojimba, T.P, (2012). Determining the effects of crude oil pollution on crop production using stochastic translog production function in Rivers State, Nigeria.*Journal of Development and Agricultural Economics* 4(13): 346-360.

Oluwatayo, I. B. & Omowunmi, T. & Ayodeji O. Land Acquisition and Use in Nigeria: Implications for Sustainable Food and Livelihood Security. *Land Use*. VALORIZA - Research Center for Endogenous Resource Valorization.

Olson, J. S. (1963). Energy Storage and the Balance of Producers and Decomposers in Ecological Systems. *Ecology* 44(2): 227-429.

Ordinioha, B. & Brisibe, S. (2013). The human health implications of crude oil spills in the Niger delta, Nigeria: An interpretation of published studies. *Nigerian Medical Journal*, 54(1):10-16.

Oshienemen, N. A. & Dilanthi, A. & Haigh, R. P. (2018). Evaluation of the Impacts of Oil Spill Disaster on Communities and Its Influence on Restiveness in Niger Delta, Nigeria Its Influence on Restiveness in Niger Delta, Nigeria. *Procedia Engineering* 212: 1054-1061.

Restuccia, D. & Da-Rocha, J.M. (2006). The Role of Agriculture in Aggregate Business Cycles. *Review of Economic Dynamics* 9(3):455-482.

Sala-ì-Martin, X. & Subramanian, A. (2013). Addressing the Natural Resource Curse: An

Illustration from Nigeria. *Journal of African Economies* 22(4): 570-615.

Uquetan, U.I. & Osang, J. E. & Egor, A.O. & Essoka, P. A. & Alozie, S.I. & Bawan, A. M. (2017). A case study of the effects of oil pollution on soil properties and growth of tree crops in Cross River State, Nigeria. *International Research Journal of Pure and Applied Physics* 5(2): 19-28.

United Nations Environmental Programme (2006): Global partnership for development: *United Nations Environmental Programme.*

United Nations Environmental Programme (2011): Environmental assessment of Ogoniland, Nairobi, Kenya: *United Nations Environmental Programme.*

U.S. Energy Information Administration (2016): Country analysis brief: Nigeria.

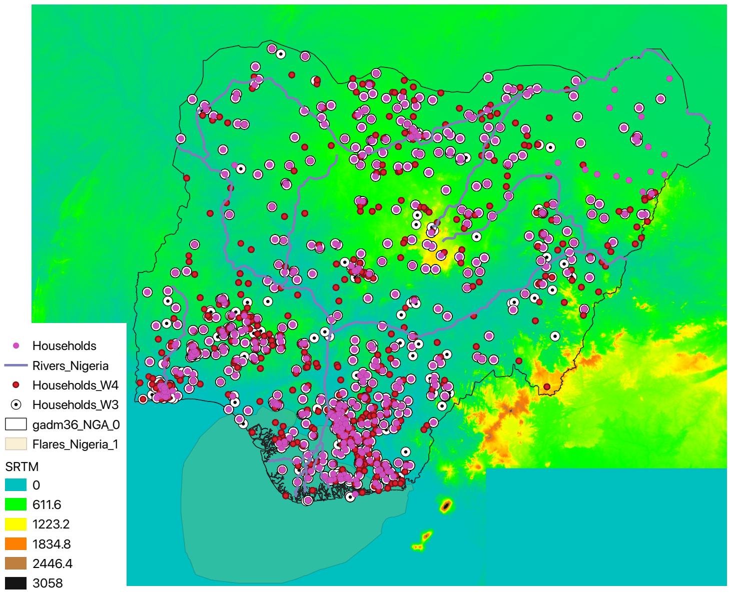
Wang Y. & Feng Jiang, L. Q. & Lyu Xianguo, W. X. & Wang G., 2013. Effects of crude oil contamination on soil physical and chemical properties in Momoge Wetland of China. Chinese Geographical Science, 23(6): 708-715.

World Bank Living Standards Measurement Study (LSMS). The General Household Survey-Panel (GHS-Panel). Years: 2010-2011, 2012-2013, 2015-1016, and 2018-2019. National Bureau of Statistics Nigeria.

World Bank (2015). More, and More Productive, Jobs for Nigeria: A Profile of Work and Workers. *International Bank for Reconstruction and Development / The World Bank.*

Zabbey, N. & Kabari, S.& Onyebuchic, A. T. (2017). Remediation of contaminated lands in the Niger Delta, Nigeria: Prospects and challenges. *Science of The Total Environment* 586, 952-965.

Figure A1: Map of Nigeria showing wetlands, rivers, enumeration areas and gas flares

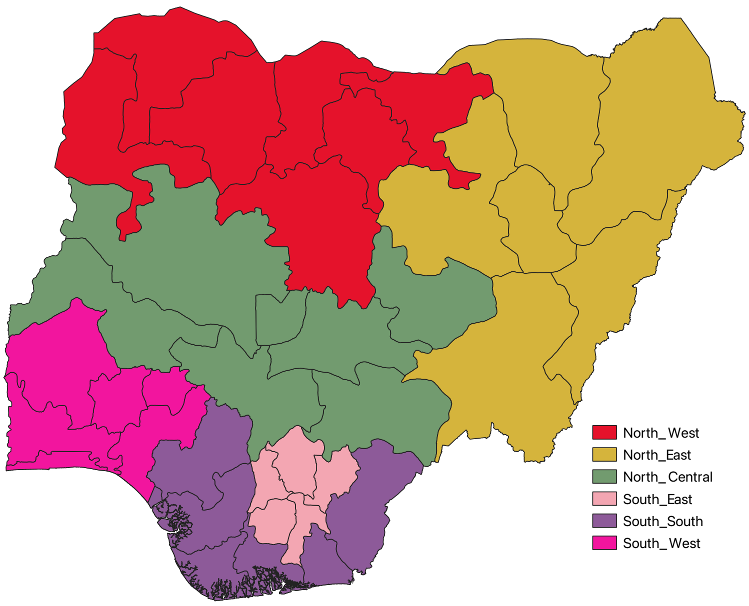


Notes: The figure shows a map of Nigeria with enumeration areas, rivers, and gas flares. Sources: Own work based on Nigerian GHS-PANEL data, gas flares shapefile from the Defense Meteorological Satellite Program (NOAA National Centers for Environmental Information), and wetlands data from Tropical and Subtropical Peatland Distribution (Center for International Forestry Research (CIFOR)).

8 Appendix

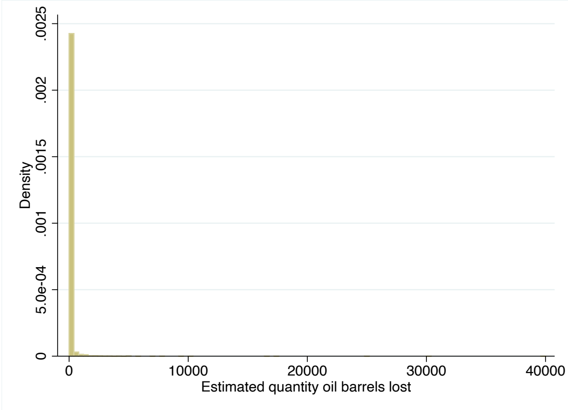
9 Figures

Figure A2: Geopolitical Zones of Nigeria

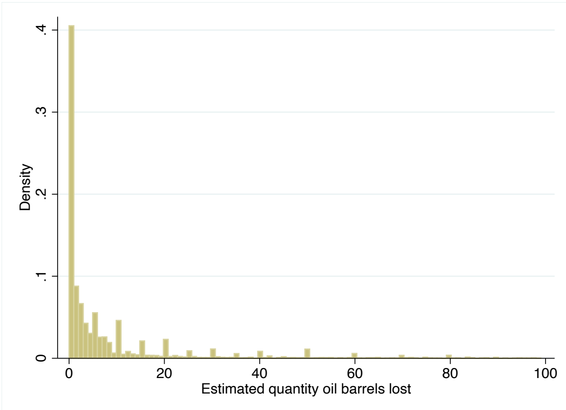


Sources: Own work based on data from the Nigerian Oil Spill Monitor (NOSDRA) and the Database of Global Administrative Areas (GADM) map for Nigeria.

Figure A3: Density function of the estimated quantity of oil lost in spills (in barrels)



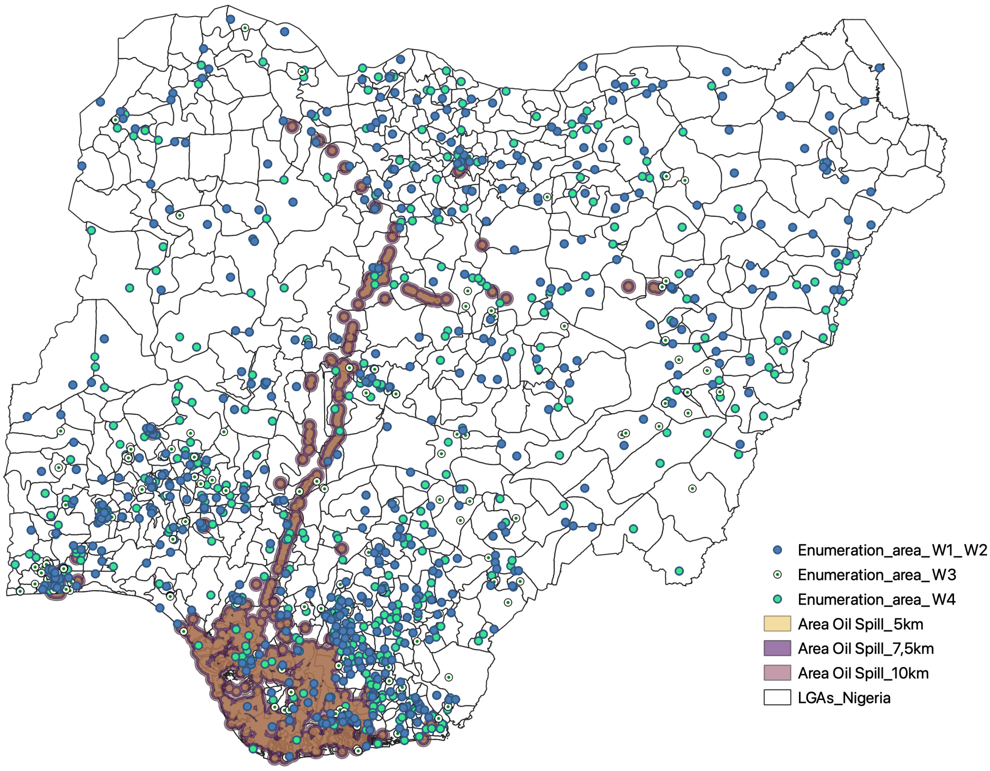
(a) Estimated quantity lost (in barrels)



(b) Estimated quantity of oil lost (in barrels) (less than 100 barrels)

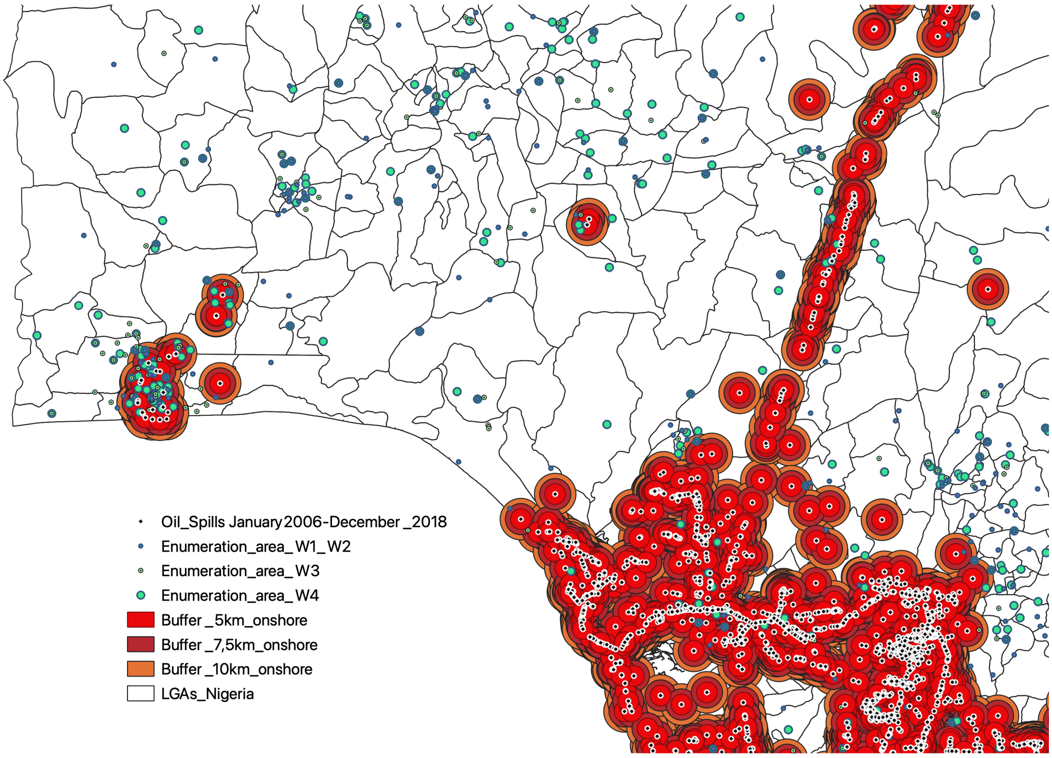
Notes: The figures show the density function for the variable âestimated quantity of oil lost in spills (in barrels)â from the table reported in the data from the Nigerian Oil Spill Monitor *NOSDRA*. Panel a) shows the density function for all the data reported, and Panel b) shows that for data reported with less than 100 barrels.

Figure A4: Map of Areas 5km to 10km from Oil-Spills



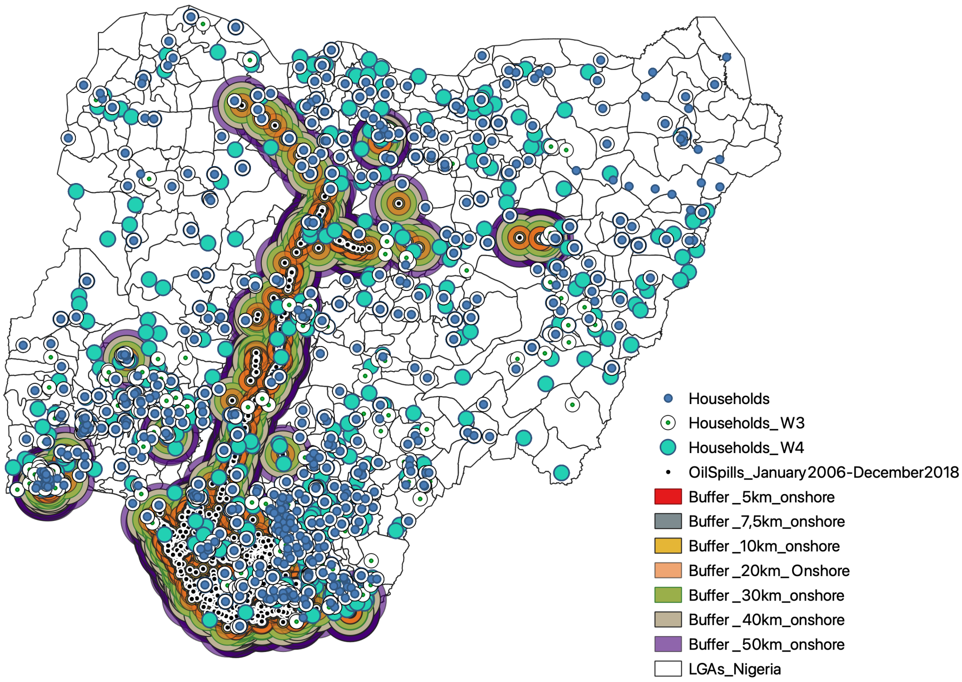
Sources: Own work based on Nigerian GHS-PANEL data and NigerianOil Spill Monitor *NOSDRA* data.

Figure A5: Zoom of locations of onshore oil spills, enumeration areas, and buffer zones at 5km, 7.5km and 10km



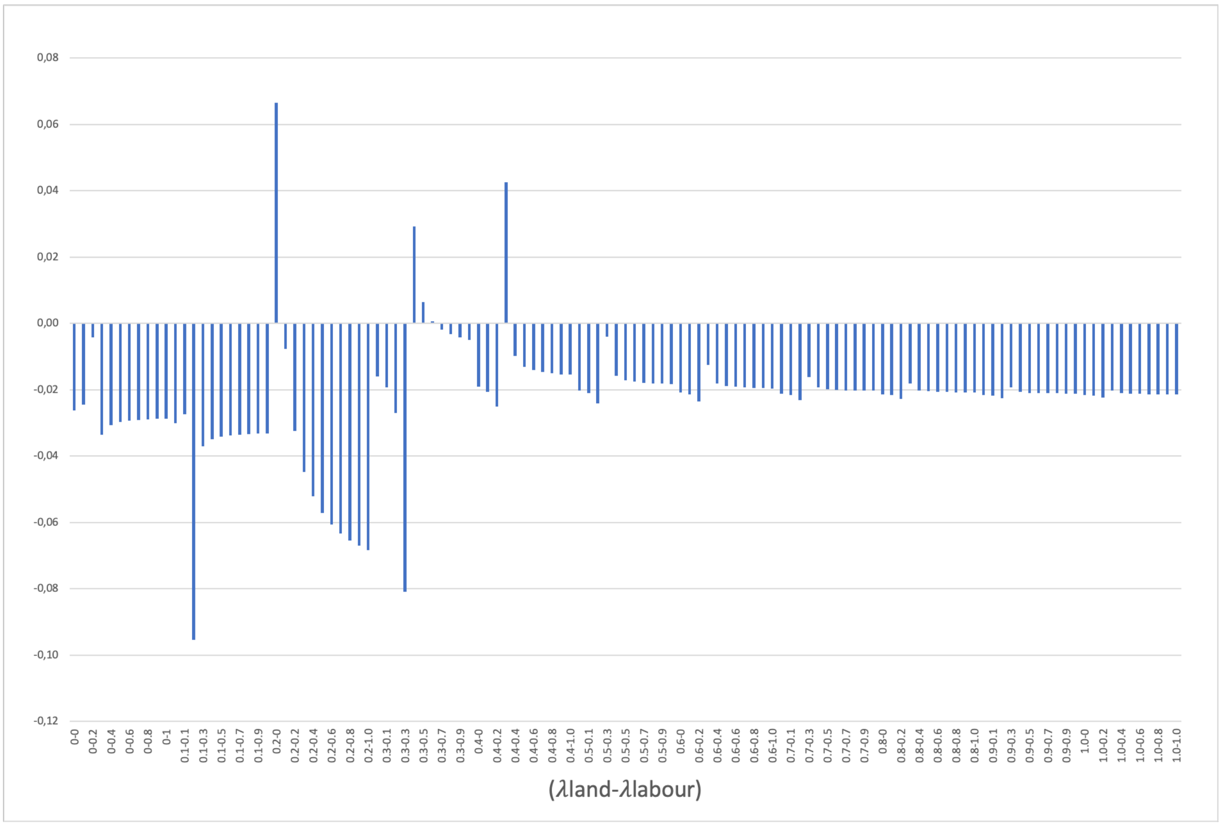
Sources: Own work based on Nigerian GHS-PANEL data and Nigerian Oil Spill Monitor *NOSDRA* data.

Figure A6: Onshore oil spills, enumeration areas, and buffers from 5km to 50km



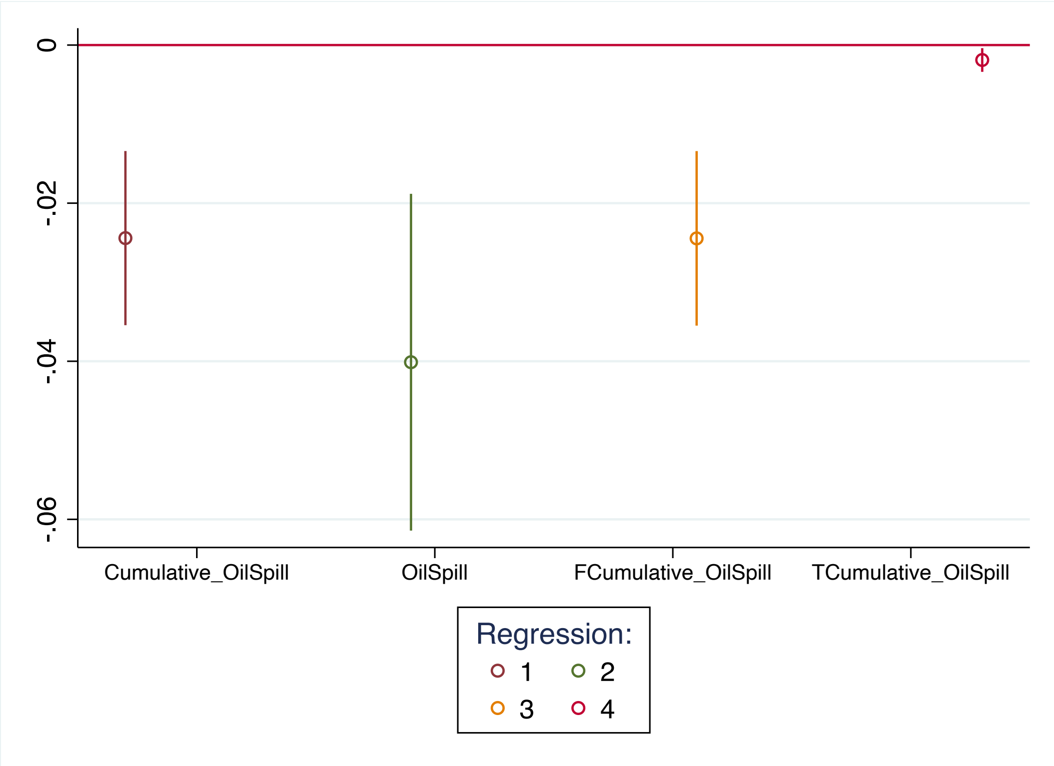
Sources: Own work based on Nigerian GHS-PANEL data and Nigerian Oil Spill Monitor*NOSDRA* data.

Figure A7: Estimates of δ with IIV approach.



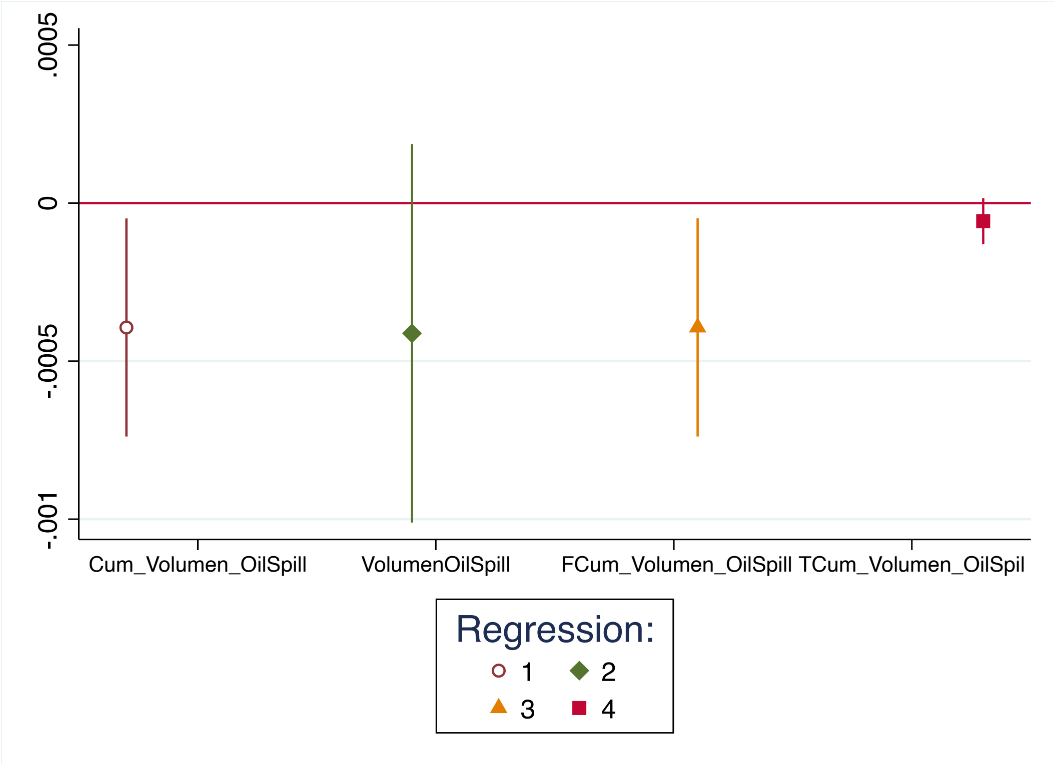
Notes: This figure shows the estimates of δ with the IIV approach.Vertical axis shows the value of δ for different values of $\lambda_{land\_harvested}$ and λ_{Labor}. Under the IIV methodology, I identify parameter bounds rather than points estimated. Each parameter measures the ratioof correlation of the instrumental variable and the regressor with the error term, which measures how well the instrument satisfies the exogeneity assumption. For instance, $\lambda_j = corr(Z_{j,\varepsilon})/corr(X_{j,\varepsilon})$ where $j = (land\_harvested, Labor)$, X is the input used, Z is the instrumental variable, and ε is the error term. The instrument is considered to be less correlated to the error term than the endogenous variable when $\varepsilon < 1$. I find that in 96% of all combinations of $\lambda_{land\_harvested}$ and λ_{Labor}, the effect on residual productivity is negative, and for all combinations where $\lambda_{land\_harvested} > 0.4$ and $\lambda_{Labor} > 0.2$, the corresponding estimate of the effect of oil-spill pollution on agricultural output is negative.

Figure A8: Effects of oil spills pollution on agricultural output. Different oil-spill pollution measurements.



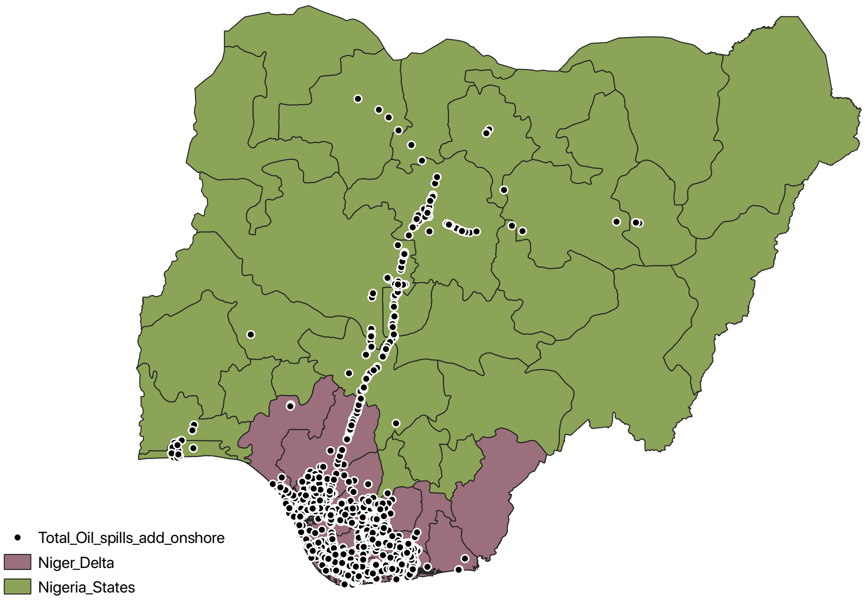
Notes: This figure shows the estimates of regressions for the effect of oil-spill pollution on agricultural output when alternative specifications are considered for the function $g(n)$. (Cumulative_OilSpills) is the variable used in the main analysis. (OilSpill) is the total number of oil-spill events per location during the year of each wave. (FCumulative_OilSpill) akes the same approach as the main analysis but I only consider the number of spills up to five years before as a persistent effect for inclusion in the exponential decay. (TCum_Volumen_OilSpil) is the total number of oil-spill events per location and wave, with no degradation effect. Circles represent point estimates, while lines indicate the 95 percent confidence interval.

Figure A9: Effects of oil-spill pollution on agricultural output. Different measurements of oil pollution. Oil pollution is approximated by the estimated number of barrels of oil lost.



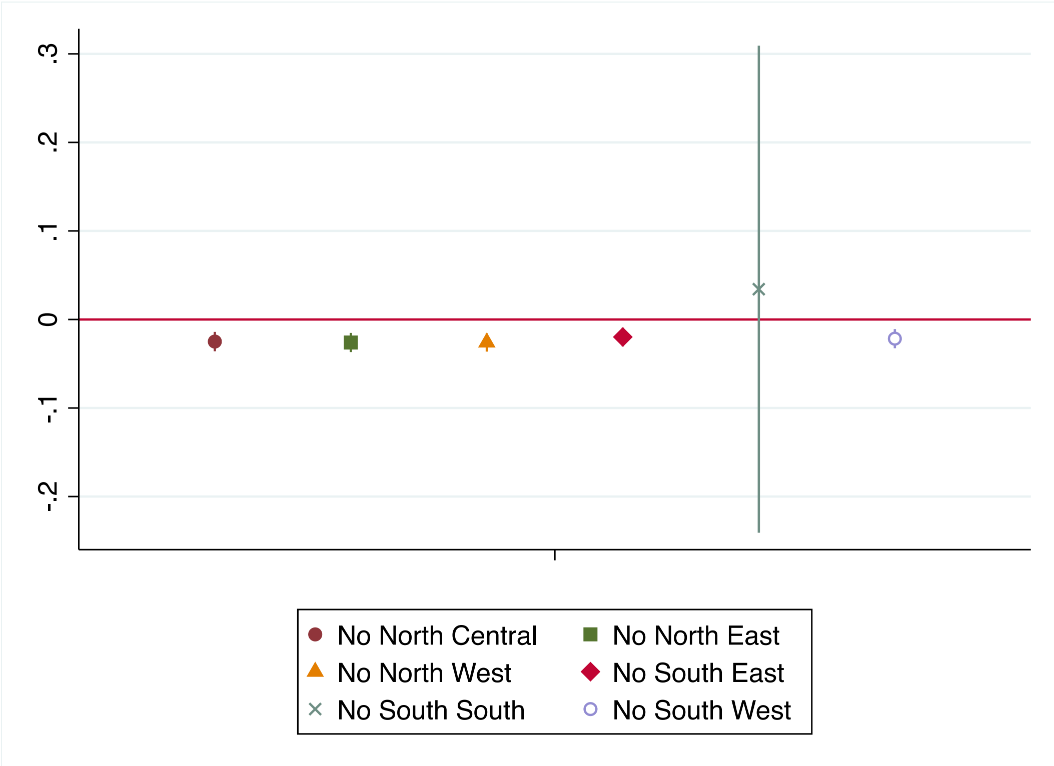
Notes: This figure shows the estimates of regressions for the effect of oil-spill pollution on agricultural output when alternative specifications are considered for the function $g(n)$. (Cum_Volumen_OilSpill barrels) is the variable used in the main robustness analysis. (Volumen_OilSpill barrels) measures the estimated number of barrels lost near a location in the year of each wave. (FCum_Volumen_OilSpill (barrels) is the total estimated number of barrels spilt near locations until the last day of the harvest survey for each wave, with no degradation effect. Geometric figures represent point estimates, while lines indicate the 95 percent confidence interval.

Figure A10: States and the Niger Delta



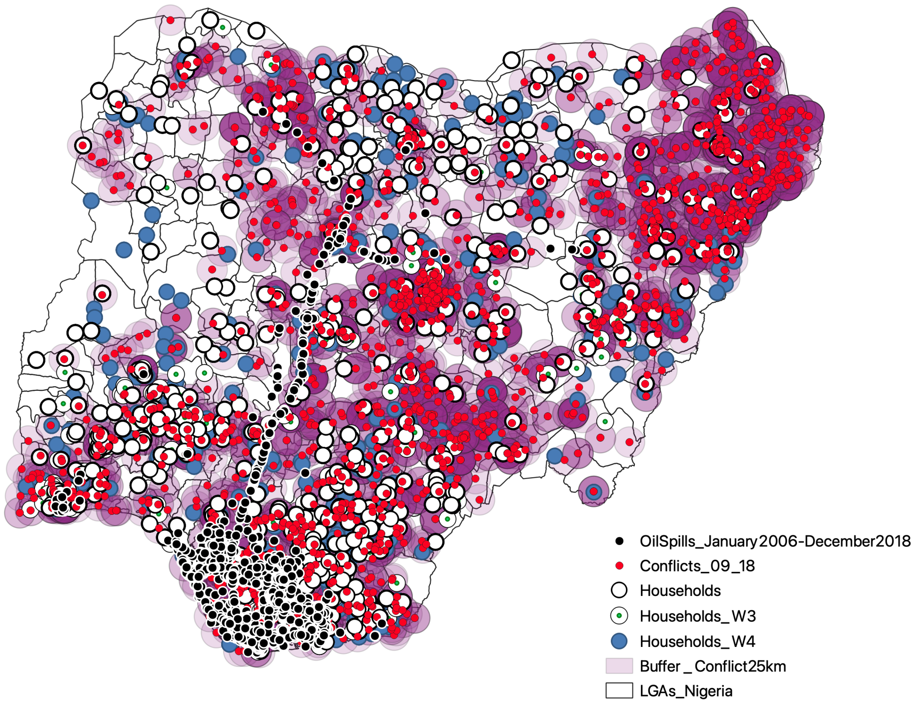
Sources: Own work based on data from the Nigerian Oil Spill Monitor *NOSDRA* and the Database of Global Administrative Areas (GADM) map for Nigeria.

Figure A11: Main specification excluding geopolitical zones



Notes: This figure shows the estimates of regressions for the effect of oil-spill pollution on agricultural output when geographical zones are dropped one by one. Geometric figures represent point estimates, while lines indicate the 95 percent confidence interval.

Figure A12: Conflicts and Buffer zones of 25 km around them. ACLED database



Notes: FThe figure shows buffer zones of 25 km around violent conflicts. Years: 2009-2018. Source: Own work based on the ACLED database.

10 Descriptive Statistics

Table B1: Nigerian GHS-PANEL (2010-2018) Main variables

Variable	Obs	Mean	Std. Dev.
Agricultural characteristics			
Real_Agricultural_Output	6948	121853.8	1424141
Land_area	6881	3.776	39.699
Labor_days	6936	403.274	683.606
Total_area_plot	6934	4.054	144.296
Endowment_Labor	6946	3.709	2.12
Harvest_Cassava	6112	461.996	6036.478
Harvest_Maize	6112	320.691	1224.193
Land_Harvested_Cassava	6112	0.278	11.275
Land_Harvested_Maize	6112	0.492	10.841
Fertility	6904	0.474	0.499
Pesticide	6932	0.232	0.422
Herbicide	6932	0.268	0.443
Improved_Seeds	6877	0.980	0.140
Yield_Maize	2107	9268.011	122933.2
Yield_Cassava	1502	7027.574	25916.07
Yield	6649	11390.3	167277.5
Household characteristics			
Age_head_household	6946	51.232	14.997
literacy	6592	0.595	0.491
Migration	6363	0.689	1.369
household_size	6948	7.683	3.995
Weather and terrain conditions characteristics			
Temperature_mean	6948	26.314	0.955
Rainfall_mean	6948	1264.633	409.651
Slope	6948	3.224	3.266
Elevation	6948	298.059	213.416
PotentialWetness	6948	14.19	3.152
Nutrient_Availability	6948	1.824	0.815
Nutrient_Retention	6948	1.574	0.602
Rooting	6948	1.415	0.719
Oxygen_to roots	6948	1.197	0.501
Exces salt	6948	1.024	0.273
Toxicity	6948	1.008	0.217
Workability	6948	1.482	0.747

Notes: Real Agricultural Output measured in 2010 Naira. Land is measured in hectares. Temperature is measured in degrees Celsius. Sample restricted to farming households. Own work using data from the Nigeria GHS-Panel.

Table B2: Nigerian GHS-PANEL (2010-2018). Additional variables

Variable	Obs	Mean	Std. Dev.
Sex	117888	1.506	.5
Age	110167	23.908	21.478
Industry	33245	3.898	4.086
AnyWork	61117	0.555	0.497
Work (agriculture)	61117	0.286	0.452
Dummy for migration	121690	0.365	0.481
Prime age male	121690	0.158	0.364
Literacy	98601	0.633	0.482
Secondary	75105	0.185	0.389
Total_hours	31887	40.551	21.897
Real_Employment_Income	10216	56175.91	591485.5
Illness_	61384	0.097	0.296
Number_days_Ill	40121	0.710	3.196
Rural	223356	0.712	0.453
Urban	222996	0.287	0.452

Notes: Full Sample. Own work using data from the Nigeria GHS-Panel.

Table B3: Summary Functions for Oil-Spill Pollution

Variable	Obs	Mean	Std. Dev.
Cumulative_Oilspill	6948	0.388	3.139
FCumulative_Oilspill	6948	0.388	3.137
Oil Spill	6948	0.238	1.987
TCumulative_Oilspill	6948	2.276	17.985
Dummy_Oil5Km	6948	0.063	0.243
Cum_Volumen_OilSpill	6854	37.803	360.02
Volumen_OilSpill	6854	37.793	359.989
FCum_Volumen_OilSpill	6697	10.035	99.128
TCum_Volumen_OilSpill	6854	216.577	2068.179

Summarize Oil pollution variables for locations within 5 km. Own work using data from NOSDRA.

11 Additional Tables

Table B4: Data Collection Dates for Surveys and Oil-Spill Incidents

		Post-Planting		Post-Harvest		Oil-spill_Incident	
Wave	Survey	Start	End	Start	End	Start	End
1	2010-2011	31/8/10	15/10/10	1/2/11	1/4/11	1/1/06	31/3/11
2	2012-2013	1/9/12	1/11/12	1/2/13	1/4/13	1/1/06	31/3/13
3	2015-2016	1/9/15	1/11/15	1/2/16	1/4/16	1/1/06	31/3/16
4	2018-2019	1/7/18	1/9/18	1/1/19	28/2/19	1/1/06	31/12/18

Notes: The table shows the data collection dates from the Nigeria GHS-surveys, and the dates used on the main variable of proxy for Oil Spill Pollution, from The Oil Spill Monitor. Note that the agriculture data on the post-harvest period of wave 4 ends in December 2018.

Table B5: First Stage Regression. Column 3 of Table 2

VARIABLES	LnLand (1)	LnLabor_days (2)
Ln(Total Own land)	0.3784***	0.1824***
	(0.020)	(0.016)
Ln(Number of adult equivalents)	0.1311***	0.4386***
	(0.036)	(0.034)
Observations	9689	10190
F-test excluded instruments	63.18***	159.21****
R-squared	0.640	0.470

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 5 kilometers of an oil spill and the interaction between that dummy and the trend. Farmer controls are included. See Table 2 for more details. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1

Table B6: Imperfect Instruments with Multiple Endogenous Variables

λ_{land}	λ_{Labor}	$\hat{\delta}$	$\hat{\alpha}$	
0	0	-0.03	0.50	0.33
0	0.1	-0.02	0.35	0.53
0	0.2	0.00	-1.39	2.86
0	0.3	-0.03	1.14	-0.53
0	0.4	-0.03	0.88	-0.18
0	0.5	-0.03	0.81	-0.08
0	0.6	-0.03	0.77	-0.04
0	0.7	-0.03	0.75	-0.01
0	0.8	-0.03	0.74	0.01
0	0.9	-0.03	0.73	0.02
0	1	-0.03	0.72	0.03
0.1	0	-0.03	0.82	0.20
0.1	0.1	-0.03	0.59	0.40
0.1	0.2	-0.10	6.27	-4.76
0.1	0.3	-0.04	1.39	-0.33
0.1	0.4	-0.03	1.22	-0.16
0.1	0.5	-0.03	1.15	-0.11
0.1	0.6	-0.03	1.12	-0.08
0.1	0.7	-0.03	1.10	-0.06
0.1	0.8	-0.03	1.09	-0.05
0.1	0.9	-0.03	1.08	-0.04
0.1	1	-0.03	1.07	-0.04
0.2	0	0.07	-7.00	3.48
0.2	0.1	-0.01	-1.02	1.25
0.2	0.2	-0.03	0.98	0.51
0.2	0.3	-0.04	1.97	0.13
0.2	0.4	-0.05	2.57	-0.09
0.2	0.5	-0.06	2.97	-0.24
0.2	0.6	-0.06	3.26	-0.35
0.2	0.7	-0.06	3.47	-0.43
0.2	0.8	-0.07	3.64	-0.49
0.2	0.9	-0.07	3.77	-0.54
0.2	1	-0.07	3.88	-0.58
0.3	0	-0.02	-0.32	0.68
0.3	0.1	-0.02	-0.07	0.75
0.3	0.2	-0.03	0.53	0.95
0.3	0.3	-0.08	4.67	2.27
0.3	0.4	0.03	-3.80	-0.44
0.3	0.5	0.01	-2.04	0.12
0.3	0.6	0.00	-1.61	0.26
0.3	0.7	0.00	-1.41	0.33
0.3	0.8	0.00	-1.30	0.36
0.3	0.9	0.00	-1.23	0.39
0.3	1	0.00	-1.18	0.40

Table B6: Continued

λ_{land}	λ_{Labor}	$\hat{\delta}$	$\hat{\alpha}$	
0.4	0	-0.02	-0.07	0.57
0.4	0.1	-0.02	0.04	0.70
0.4	0.2	-0.03	0.37	1.11
0.4	0.3	0.04	-4.55	-5.04
0.4	0.4	-0.01	-0.74	-0.27
0.4	0.5	-0.01	-0.51	0.01
0.4	0.6	-0.01	-0.43	0.12
0.4	0.7	-0.01	-0.39	0.17
0.4	0.8	-0.02	-0.36	0.20
0.4	0.9	-0.02	-0.34	0.22
0.4	1	-0.02	-0.33	0.24
0.5	0	-0.02	0.02	0.53
0.5	0.1	-0.02	0.08	0.67
0.5	0.2	-0.02	0.28	1.20
0.5	0.3	0.00	-1.08	-2.28
0.5	0.4	-0.02	-0.28	-0.25
0.5	0.5	-0.02	-0.19	-0.01
0.5	0.6	-0.02	-0.15	0.08
0.5	0.7	-0.02	-0.13	0.13
0.5	0.8	-0.02	-0.12	0.16
0.5	0.9	-0.02	-0.11	0.18
0.5	1	-0.02	-0.11	0.19
0.6	0	-0.02	0.07	0.51
0.6	0.1	-0.02	0.10	0.66
0.6	0.2	-0.02	0.23	1.25
0.6	0.3	-0.01	-0.43	-1.78
0.6	0.4	-0.02	-0.10	-0.24
0.6	0.5	-0.02	-0.05	-0.02
0.6	0.6	-0.02	-0.03	0.07
0.6	0.7	-0.02	-0.02	0.11
0.6	0.8	-0.02	-0.01	0.14
0.6	0.9	-0.02	-0.01	0.16
0.6	1	-0.02	-0.01	0.17
0.7	0	-0.02	0.10	0.50
0.7	0.1	-0.02	0.11	0.65
0.7	0.2	-0.02	0.19	1.28
0.7	0.3	-0.02	-0.16	-1.56
0.7	0.4	-0.02	0.00	-0.23
0.7	0.5	-0.02	0.03	-0.03
0.7	0.6	-0.02	0.04	0.06
0.7	0.7	-0.02	0.04	0.10
0.7	0.8	-0.02	0.05	0.13
0.7	0.9	-0.02	0.05	0.15
0.7	1	-0.02	0.05	0.16

Table B6: Continued

λ_{land}	λ_{Labor}	$\hat{\delta}$	$\hat{\alpha}$	
0.8	0	-0.02	0.11	0.49
0.8	0.1	-0.02	0.12	0.65
0.8	0.2	-0.02	0.17	1.31
0.8	0.3	-0.02	-0.02	-1.44
0.8	0.4	-0.02	0.07	-0.23
0.8	0.5	-0.02	0.08	-0.03
0.8	0.6	-0.02	0.08	0.05
0.8	0.7	-0.02	0.09	0.10
0.8	0.8	-0.02	0.09	0.12
0.8	0.9	-0.02	0.09	0.14
0.8	1	-0.02	0.09	0.16
0.9	0	-0.02	0.13	0.49
0.9	0.1	-0.02	0.13	0.65
0.9	0.2	-0.02	0.15	1.33
0.9	0.3	-0.02	0.08	-1.37
0.9	0.4	-0.02	0.11	-0.22
0.9	0.5	-0.02	0.11	-0.03
0.9	0.6	-0.02	0.12	0.05
0.9	0.7	-0.02	0.12	0.09
0.9	0.8	-0.02	0.12	0.12
0.9	0.9	-0.02	0.12	0.14
0.9	1	-0.02	0.12	0.15
1	0	-0.02	0.14	0.48
1	0.1	-0.02	0.14	0.64
1	0.2	-0.02	0.13	1.34
1	0.3	-0.02	0.14	-1.32
1	0.4	-0.02	0.14	-0.22
1	0.5	-0.02	0.14	-0.03
1	0.6	-0.02	0.14	0.04
1	0.7	-0.02	0.14	0.09
1	0.8	-0.02	0.14	0.11
1	0.9	-0.02	0.14	0.13
1	1	-0.02	0.14	0.15

Notes: The tables show the estimates used to construct Figure 3

Table B7: Additional checks: Spatial disaggregation

VARIABLES	Ln Agricultural Output					
	(1)	(2)	(3)	(4)	(5)	(6)
Spills_5	-0.0328*	-0.0335	-0.0383	-0.0390*	-0.0395*	-0.0373*
	(0.018)	(0.021)	(0.025)	(0.021)	(0.021)	(0.022)
Spills_7	-0.0822	-0.0864	-0.1267	-0.1040	-0.0799	-0.0584
	(0.107)	(0.130)	(0.126)	(0.130)	(0.138)	(0.134)
Spills_10	-0.0066	-0.0085	-0.0096	-0.0124	-0.0116	-0.0107
	(0.014)	(0.017)	(0.020)	(0.017)	(0.017)	(0.018)
Spills_20	0.0096	0.0062	0.0038	0.0039	0.0039	0.0028
	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)
Spills_30	0.0067	0.0119	0.0291	0.0089	0.0087	0.0139
	(0.076)	(0.061)	(0.070)	(0.063)	(0.064)	(0.063)
Spills_40	-0.0027	-0.0018	-0.0011	-0.0014	-0.0014	-0.0014
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Spills_50	0.0109**	0.0116**	0.0126**	0.0120**	0.0119**	0.0120**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
LnLand		0.2577***	0.5403***	0.2474***	0.2467***	0.2471***
		(0.023)	(0.171)	(0.023)	(0.023)	(0.023)
LnLabor_days		0.2010***	0.2940*	0.1915***	0.1918***	0.1883***
		(0.024)	(0.153)	(0.024)	(0.024)	(0.023)
Estimation	OLS	OLS	2SLS	OLS	OLS	OLS
Observations	6369	6115	6114	6059	6059	6059
R-squared	0.610	0.640	0.619	0.643	0.643	0.644
Farmer controls	YES	YES	YES	YES	YES	YES
Other inputs	NON	NON	NON	YES	YES	YES
Heterogeneous trend	NON	NON	NON	NON	YES	YES
Climate & environmental variables	NON	NON	NON	NON	NON	YES
Waves dummies	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors clustered at the LGAs levels. Spill_5 to Spill_30 are the total number of Oil spills in with an exponential decay over time, at distance of 5 km to 30 km from nearby locality e. Farmers controls: age of head of household and literacy, an indicator of whether the household owns its farm plot, Heterogenous trend: distances to federal road, main towns, main markets, states capitals, and border post on the main road. Climate and environmental variables: mean temperature and rainfall, rooting, slope, nutrient retention, excel salt, oxygen to roots, toxicity, and workability. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1 Columns 3 is estimated using 2SLS. The excluded instruments are the log of the area of land managed, and the log of the number of adults equivalents in the household.

Table B8: Additional checks: Main Results in enumeration areas up to 7.5km from oil spills

VARIABLES	LnAg._Output	LnAg._Output	LnAg._Output	LnYield	LnYield_Cassava	LnYield_Maize
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative_OilSpill	-0.0307	-0.0248*	-0.0241**	-0.0157***	-0.0210**	-0.0047
	(0.027)	(0.015)	(0.010)	(0.006)	(0.009)	(0.011)
LnLand		0.2577***	0.5373***			
		(0.023)	(0.171)			
LnLabor_days		0.1992***	0.2967*			
		(0.024)	(0.153)			
Estimation	OLS	OLS	2SLS	OLS	OLS	OLS
Observations	6342	5998	5935	9056	1101	2493
R-squared	0.612	0.632	0.604	0.312	0.293	0.290
Waves dummies	YES	YES	YES	YES	YES	YES
LGAs fixed effects	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 7.5 kilometers of an oil spill. Farmer controls: columns 2 to 4 and column 6 include age of head of household, literacy, and an indicator of whether the household owns its farm plot. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1 Column 3 is estimated using 2SLS. The excluded instruments are the log of the area of land managed and the log of the number of equivalent adults in the household.

Table B9: Additional checks: Additional distance robustness in enumeration areas near oil spills

Variable	Ln Agricultural Output									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cumulative_OilSpill_10km	0.0046	0.0023								
	(0.011)	(0.014)								
Cumulative_OilSpill_20km			0.0055	0.0028						
			(0.011)	(0.011)						
Cumulative_OilSpill_30km					-0.0114**	-0.0119***				
					(0.005)	(0.004)				
Cumulative_OilSpill_40km							-0.0014	-0.0008		
							(0.002)	(0.002)		
Cumulative_OilSpill_50km									0.0012	0.0024
									(0.003)	(0.003)
Lntmp_area	0.2572***	0.5222***	0.2569***	0.5280***	0.2574***	0.5052***	0.2565***	0.5174***	0.2575***	0.5162***
	(0.024)	(0.175)	(0.024)	(0.173)	(0.024)	(0.176)	(0.024)	(0.176)	(0.024)	(0.176)
LnLabor_days	0.2007***	0.3225**	0.2011***	0.3206**	0.2006***	0.3160*	0.2010***	0.3261**	0.2006***	0.3240**
	(0.025)	(0.162)	(0.025)	(0.161)	(0.025)	(0.162)	(0.025)	(0.162)	(0.025)	(0.163)
Estimation	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Observations	6130	6114	6130	6114	6130	6114	6130	6114	6130	6114
R-squared	0.640	0.016	0.640	0.014	0.641	0.025	0.640	0.017	0.640	0.017
Waves dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA levels. All regressions include a dummy for being within 10, 20, 30, 40 or 50 kilometers, depending on the distance considered to construct the Cumulative OilSpill variable. Farmer controls: columns 1 to 10 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Significance is denoted as follows: *** $p<0.01$. ** $p<0.05$. * $p<0.1$.

Table B10: Additional checks: 2SLS Estimation

VARIABLES	Ln Agricultural Output			
	(1)	(2)	(3)	(4)
Cumulative_OilSpill	-0.0276***	-0.0284***	-0.0257***	-0.0264***
	(0.007)	(0.007)	(0.007)	(0.007)
LnLand	0.5356***	0.5494***	0.5780***	0.5933***
	(0.177)	(0.179)	(0.184)	(0.191)
LnLabor_days	0.2717*	0.2636*	0.2690*	0.2332
	(0.155)	(0.156)	(0.162)	(0.165)
Fertilizers	0.2129***	0.2173***		0.2254***
	(0.073)	(0.072)		(0.072)
Pesticides	0.1313*	0.1290*		0.1189
	(0.073)	(0.073)		(0.074)
Herbicides	0.0135	0.0202		0.0172
	(0.097)	(0.099)		(0.100)
Improved_Seeds	-0.0290	-0.0172		-0.0050
	(0.271)	(0.272)		(0.272)
Rooting			0.4527**	0.4519**
			(0.217)	(0.213)
Oxygen to roots			-0.0791	-0.0429
			(0.191)	(0.194)
Toxicity			0.1821	-0.0054
			(0.578)	(0.564)
Excessalt			-0.1781	-0.0875
			(0.505)	(0.484)
Workability			-0.4099**	-0.4070**
			(0.205)	(0.206)
Nutrient_Retention			0.3319	0.3698
			(0.243)	(0.242)
Nutrient_Availability			-0.2897	-0.2917
			(0.222)	(0.220)
Mean temperature			0.0003	0.0032
			(0.013)	(0.013)
Mean rainfall			0.0006	0.0007
			(0.001)	(0.001)
Slope			0.0204	0.0195
			(0.018)	(0.017)
Estimation	2SLS	2SLS	2SLS	2SLS
Observations	6058	6058	6114	6058
R-squared	0.622	0.620	0.615	0.617
Waves dummies	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at the LGAs levels. All regressions include a dummy for being within 5 kilometers of an oil spill. Farmer controls: columns 1 to 5 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Columns 3 to 5 include indicators from time trends with distances to federal roads, main towns, main markets, states capitals, and border posts on the main road. The excluded instruments are the log of the area of land managed and the log of the number of equivalent adults in the household. Significance is denoted as follows*** p<0.01. ** p<0.05. * p<0.1.

Table B11: Additional checks: Alternative measurements of oil pollution

VARIABLES	LnAg._Output (1)	LnAg._Output (2)	LnYield (3)	LnAg._Output (4)	LnAg._Output (5)	LnYield (6)	LnAg._Output (7)	LnAg._Output (8)	LnYield (9)
OilSpill	-0.0388***	-0.0440***	-0.0377						
	(0.010)	(0.013)	(0.038)						
FCumulative_OilSpill				-0.0234***	-0.0274***	-0.0285*			
				(0.005)	(0.007)	(0.017)			
TCumulative_OilSpill							-0.0017**	-0.0022**	-0.0039***
							(0.001)	(0.001)	(0.001)
LnLand	0.2572***	0.5390***		0.2572***	0.5384***		0.2572***	0.5361***	
	(0.023)	(0.172)		(0.023)	(0.171)		(0.023)	(0.171)	
LnLabor_days	0.2000***	0.2951*		0.2001***	0.2954*		0.2001***	0.2979*	
	(0.024)	(0.153)		(0.024)	(0.153)		(0.024)	(0.153)	
Estimation	OLS	2SLS	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS
Observations	6115	6114	8887	6115	6114	8887	6115	6114	8887
R-squared	0.639	0.618	0.316	0.639	0.618	0.316	0.639	0.618	0.316
Waves dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA levels. All regressions include a dummy for being within 5 kilometers of an oil spill. Total_OilSpill measures the number of oil spills near a location in the year of each wave. FCumulative_OilSpill measures cumulative oil spills persistent for five years only. TCumulative_OilSpil is the total number of oil spills at a location within 5km up to the last day of the harvest survey for each wave. Farmer controls: columns 1, 2, 4, and 5, and columns 7 and 8 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1 The excluded instruments are the log of the area of land managed and the log of the number of adults equivalents in the household.

Table B12: Additional checks: Alternative measurements of oil pollution. Estimated number of barrels lost in spills.

VARIABLES	LnAg._Output (1)	LnAg._Output (2)	LnYield (3)	LnAg._Output (4)	LnAg._Output (5)	LnYield (6)	LnAg._Output (7)	LnAg._Output (8)	LnYield (9)	LnAg._Output (10)	LnAg._Output (11)	LnYield (12)
Cum_Volumen_OilSpill (barrels)	-0.0004**	-0.0004**	-0.0000									
	(0.000)	(0.000)	(0.000)									
Volumen_OilSpill (barrels)				-0.0004	-0.0005	-0.0002						
				(0.000)	(0.000)	(0.000)						
FCum_Volumen_OilSpill (barrels)							-0.0004**	-0.0004**	-0.0000			
							(0.000)	(0.000)	(0.000)			
TCum_Volumen_OilSpil (barrels)										-0.0001	-0.0001*	-0.0001*
										(0.000)	(0.000)	(0.000)
LnLand	0.2560***	0.5364***	-0.6918***	0.2524***	0.4868***	-0.6959***	0.2560***	0.5364***		0.2562***	0.5340***	-0.6917***
	(0.024)	(0.177)	(0.019)	(0.024)	(0.187)	(0.019)	(0.024)	(0.177)		(0.024)	(0.177)	(0.019)
LnLabor_days	0.2020***	0.3012*	0.2319***	0.1973***	0.3083*	0.2340***	0.2020***	0.3012*		0.2018***	0.3035*	0.2319***
	(0.024)	(0.157)	(0.017)	(0.024)	(0.166)	(0.018)	(0.024)	(0.157)		(0.024)	(0.157)	(0.017)
Estimation	OLS	2SLS	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS
Observations	6035	6034	8709	5893	5892	8481	6035	6034	8768	6035	6034	8709
R-squared	0.634	0.613	0.530	0.636	0.620	0.532	0.634	0.613	0.323	0.634	0.613	0.530
Waves dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA levels. All regressions include a dummy for being within 5 kilometers of an oil spill. Cum_Volumen_OilSpill (barrels) is the total estimated number of barrels spilt near locations up to the last day of the harvest survey for each wave degraded over time. Volumen_OilSpill (barrels) measures the estimated number of barrels lost in spills near a location in the year of each wave. FCum_Volumen_OilSpill (barrels) measures cumulative oil spills persisting for five years only. TCum_Volumen_OilSpil (barrels) is the total estimated number of barrels spilt near locations up to the last day of the harvest survey for each wave. Farmers control: columns 1and 2. 4 and 5. and columns 7 and 8 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Significance is denoted as follows *** p<0.01. ** p<0.05. * p<0.1 The excluded instruments are the log of the area of land managed and the log of the number of equivalent adults in the household.

Table B13: Additional checks: Additional checks: Additional robustness in enumeration areas within 7.5km of oil spills

VARIABLES	LnAg._Output (1)	LnAg._Output (2)	LnAg._Output (3)	LnAg._Output (4)	LnAg._Output (5)
Cumulative_OilSpill	-0.0317	-0.0256*	-0.0280*	-0.0232*	-0.0259*
	(0.027)	(0.015)	(0.015)	(0.014)	(0.015)
LnLand		0.2471***	0.2464***	0.2571***	0.2468***
		(0.023)	(0.023)	(0.024)	(0.023)
LnLabor_days		0.1895***	0.1900***	0.1969***	0.1865***
		(0.024)	(0.024)	(0.023)	(0.023)
Fertilizers	0.3235***	0.2591***	0.2614***		0.2707***
	(0.076)	(0.073)	(0.072)		(0.072)
Pesticides	0.3397***	0.2192***	0.2151***		0.2098***
	(0.066)	(0.064)	(0.065)		(0.064)
Herbicides	0.1921**	0.0876	0.0997		0.1025
	(0.095)	(0.095)	(0.095)		(0.095)
Improved_Seeds	0.2539	0.1518	0.1608		0.1487
	(0.338)	(0.263)	(0.262)		(0.261)
Rooting				0.4520*	0.4450*
				(0.253)	(0.239)
Oxygen to roots				-0.0503	-0.0214
				(0.186)	(0.188)
Toxicity				0.6026	0.3784
				(0.492)	(0.495)
Exces salt				-0.5721	-0.4171
				(0.399)	(0.405)
Workability				-0.4471*	-0.4441*
				(0.237)	(0.229)
Nutrient_Retention				0.2577	0.2847
				(0.243)	(0.238)
Nutrient_Availability				-0.2598	-0.2581
				(0.214)	(0.210)
Mean temperature				0.0134	0.0170
				(0.014)	(0.014)
Mean rainfall				-0.0001	0.0001
				(0.000)	(0.000)
Slope				0.0126	0.0106
				(0.017)	(0.016)
Estimation	OLS	OLS	OLS	OLS	OLS
Observations	6275	5943	5943	5998	5943
R-squared	0.618	0.636	0.636	0.634	0.638
Waves dummies	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 7.5 kilometers of an oil spill. Farmer controls: Columns 1 to 5 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Columns 3 to 5 include indicators from time trends with distances to federal roads, main towns, main markets, states capitals, and border posts on the main road. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1.

Table B14: Additional checks: Specific Zones and Areas

DEPENDENT VARIABLE	LnAg._Output							
	Non NCentral	Non NEast	Non NWest	Non SEast	Non SSouth	Non SWest	Niger Delta	ND/Out ND
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cumulative_OilSpill	-0.0250***	-0.0260***	-0.0259***	-0.0198***	0.0342	-0.0217***	-0.0346***	
	(0.006)	(0.006)	(0.005)	(0.005)	(0.140)	(0.006)	(0.007)	
(A) Cumulative_OilSpill*NigerDelta								-0.8711***
								(0.269)
(B) Cumulative_OilSpill*Outside NigerDelta								-0.5180***
								(0.270)
Test (A)-(B)								
p-value								0.144
LnLand	0.2749***	0.2680***	0.2698***	0.2583***	0.2385***	0.2295***	0.2706***	0.2228***
	(0.026)	(0.026)	(0.024)	(0.028)	(0.024)	(0.025)	(0.042)	(0.028)
LnLabor_days	0.1805***	0.1997***	0.1981***	0.2197***	0.2093***	0.1925***	0.2037***	0.1641***
	(0.026)	(0.026)	(0.025)	(0.028)	(0.025)	(0.025)	(0.043)	(0.027)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Observations	4883	4954	4897	4917	5431	5487	1402	4149
R-squared	0.647	0.629	0.636	0.652	0.641	0.627	0.607	0.465
Waves dummies	YES	YES	YES	YES	YES	YES	YES	NON
Year fixed effects	NON	NON	NON	NON	NON	NON	NON	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
EA X trend	NON	NON	NON	NON	NON	NON	NON	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 5 kilometers of an oil spill. Farmer controls: Columns 1 to 8 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Columns 1 to 6 exclude one specific area. Column 8 restricts the analysis to the Niger Delta and column 9 restricts the analysis from Wave 1 to Wave 3 with panel data compare the Niger Delta and non-Niger Delta areas. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1.

Table B15: Additional checks: Conflict versus Non Conflict Areas

VARIABLES	LnAg._Output (1) CONFLICT AREA	LnAg._Output (2) NON CONFLICT AREA	LnAg._Output (3)
Cumulative_OilSpill	-0.0086*	-1.2303*	
	(0.005)	(0.630)	
(A) Cumulative_OilSpill*Conflict Area			-0.8038***
			(0.274)
(B) Cumulative_OilSpill*Non-Conflict Area			-1.1970***
			(0.221)
Test (A)-(B)			
p-value			0.017
LnLand	0.3323***	0.1686***	0.2478***
	(0.032)	(0.055)	(0.028)
LnLabor_days	0.1703***	0.1741***	0.1602***
	(0.038)	(0.044)	(0.032)
Observations	2917	1014	3310
R-squared	0.656	0.489	0.451
Waves dummies	YES	YES	NON
Year fixed effects	NON	NON	YES
LGA fixed effects	YES	YES	YES
EA X trend	NON	NON	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 5 kilometers of an oil spill. Farmer controls: Columns 1 to 3 include age of head of household, literacy, and an indicator of whether a household owns its farm plot. Significance is denoted as follows: *** p<0.01. ** p<0.05. * p<0.1.

Table B16: Oil Spill Pollution during the year and Self-reported Illness

Variable	Ill in previous four weeks			Ln (Number of days off work)		
	(1)	(2)	(3)	(4)	(5)	(6)
OilSpill	0.0004	-0.0021	0.0010**	-0.0012	0.0594	-0.0045
	(0.001)	(0.001)	(0.000)	(0.003)	(0.049)	(0.003)
Sample	All	Urban	Rural	All	Urban	Rural
Observations	36675	9498	26196	23001	5916	16491
R-squared	0.070	0.056	0.080	0.225	0.178	0.252
Waves dummies	YES	YES	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA level. All regressions include a dummy for being within 5 kilometers of an oil spill and individual controls such as age, age^2, gender, an indicator of ecological zone, and rural area. "Ill in previous four weeks" is an indicator that takes a value of 1 if any individual reports being ill during the last four weeks. This does not include injures. "Ln (Number of days off work)" is the log of the number of days that an individual reports ceasing to engage in any usual activity. Significance is denoted as follows: *** p<0.01, ** p<0.05, * p<0.1

Table B17: Oil Spill Pollution and Labor Outcomes for urban workers

Variable	Ln(Total hours worked)		Ln(Real Employment Income)	
	(1)	(2)	(3)	(4)
OilSpill	-0.0108***	-0.0084***	-0.2201***	-0.2275***
	(0.004)	(0.003)	(0.012)	(0.012)
Sample	All urban workers	Urban non-agric. workers	All urban workers	Urban non-agric. workers
Observations	4782	4369	2846	2788
R-squared	0.240	0.130	0.398	0.389
Waves dummies	YES	YES	YES	YES
LGA fixed effects	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses. Standard errors are clustered at LGA levels. All regressions include a dummy for being within 5 kilometers of an oil spill, and an industrial dummy. Columns 1 and 2 include individual controls such as age, age^2, literacy status, and household size. Columns 3 and 4 add additional controls in the form of the log of the total number of hours worked. All regressions exclude oil industry workers. Significance is denoted as follows: *** p<0.01, ** p<0.05, * p<0.1