Winning the Oil Lottery: The Impact of Natural Resource Extraction on Growth*

Tiago Cavalcanti† Daniel Da Mata‡ Frederik Toscani§

Version: January 27, 2015.

Abstract

This paper provides evidence of the causal impact of oil discoveries on local development. Novel data on the drilling of 20,000 oil wells in Brazil allows us to exploit a quasi-experiment: Municipalities where oil was discovered constitute the treatment group, while municipalities with drilling but no discovery are the control group. The results show that oil discoveries significantly increase per capita GDP and urbanization. We find positive spillovers to non-oil sectors, specifically, an increase in services GDP which stems from higher output per worker. The results are consistent with greater local demand for non-tradable services driven by highly paid oil workers.

Keywords: Oil and Gas, Economic Growth, Urbanization.

JEL: O13, O40

*We have benefited from discussions with Toke Aidt, Rabah Arezki, Francesco Caselli, Silvio Costa, Claudio Ferraz, Jane Fruehwirth, Sriya Iyer, Hamish Low, Kaivan Munshi, Sheilagh Ogilvie, Andre Pereira, Pontus Rendahl, Cezar Santos, Rodrigo Serra, Edson Severnini, Claudio Souza, Daniel Sturm, José Tavares and Jaume Ventura. We thank the Keynes Fund, University of Cambridge, for financial support. All remaining errors are ours. The views expressed in this article are those of the author(s) and do not necessarily represent those of the IMF, IMF policy, or IPEA.

†Faculty of Economics, University of Cambridge.

‡Institute for Applied Economic Research (IPEA).

§International Monetary Fund (IMF).

"No other business so starkly and extremely defines the meaning of risk and reward — and the profound impact of chance and fate." *Yergin (2008)*

1 Introduction

What are the effects of oil discoveries on economic development? Although there is a long tradition in economics of studying the impact of natural resource abundance, no clear consensus has emerged in the literature. Should the discovery of oil lead to a prosperous period of high growth in both the short and long run or should countries fear the much-discussed Dutch disease? Competing theories and empirical evidence point out that the pure market effect of natural resource extraction might drive up local prices — and therefore crowd out the development of other economic activities, bringing about negative effects on growth — or increase demand for workers and attract new activities, which can lead to agglomeration effects, with a positive impact on productivity and income.

This paper uses the quasi-experiment generated by the random outcomes of exploratory oil drilling in Brazil in order to investigate the causal effect of natural resource discoveries on local development.[1] Specifically, we compare economic outcomes in municipalities where the national oil company, Petrobras, drilled for oil but did not find any, to outcomes in those municipalities in which it drilled for oil and was successful.[2] Drilling attempts were carried out in many locations with similar geological characteristics, but oil was found in only a few places. The "treatment assignment" is related to the success of drilling attempts: Places where oil was found were assigned to treatment, while places with no oil are part of the control group. The treatment assignment resembles a "ran-

[1] Oil and gas are also called petroleum or hydrocarbons. Throughout this paper, we use the term *oil* to refer to *oil and gas*. The oil industry is loosely divided into two segments: upstream and downstream. Upstream refers to exploration and production of oil, while downstream refers to processing and transportation (refineries, terminals, etc).

[2] There are three administrative levels in Brazil: federal government, states, and municipalities. Municipalities are autonomous entities that are able, for instance, to set property and service taxes. They are roughly equivalent to counties in the U.S. We use the terms *municipality*, *local government* and *local economy* interchangeably.

domization", since (conditional on drilling taking place) a discovery depends mainly on luck. Therefore, places with oil discoveries are the "winners" of the "geological lottery." Since there were no significant royalty payments to municipalities in Brazil until several decades after the first discoveries, we are able to isolate the *direct impact* of oil extraction from the effect of fiscal windfalls.

Our analysis uses novel data on the drilling of approximately 20,000 oil wells in Brazil from 1940 to 2000. The dataset covers the universe of wells drilled since exploration began in the country and provides information on three stages regarding oil extraction and production: drilling, discovery, and upstream production. We use this detailed information to distinguish those municipalities which were assigned to treatment from those which constitute the control group. Since we view production as the treatment, and discovery as the assignment to treatment, our focus is on an Intent-to-Treat (ITT) analysis, where we regress our outcome variables of interest directly on discoveries.[3] Discoveries take place in different locations over time, so we can exploit time and cross-sectional variations. The ITT analysis enables us to obtain a lower bound on the average treatment effect. We also estimate a Local Average Treatment Effect (LATE) by instrumenting for production with discoveries.[4] In addition, we study treatment intensity using detailed information on different types of wells. This allows us to retrieve a coefficient that can be interpreted as a weighted average of per-unit treatment effect.

The baseline results show that locations in which oil was discovered had a 24.6–25.9% higher *per capita* GDP over a span of up to 60 years compared to those in the control group. Furthermore, we document an increase in both manufacturing and services GDP *per capita* but no impact on agricultural GDP. While the measure of manufacturing GDP includes natural resource extraction (and as such an increase is not surprising), the increase in services indicates spillover effects of oil production impacting the rest of the economy. Additionally, we find evidence for an increase in urbanization of about 4 percentage points. This increase in urbanization is consistent with the increase in services

[3]Some municipalities discover oil but do not extract it.

[4]Endogeneity of production might be more of a problem for gas than for oil. While it is relatively easy to transport oil, gas requires a substantial investment in infrastructure such as pipelines.

we document. We do not find any effect on population density. Using historical data on sectoral employment, we calculate a measure of sectoral output per worker and show that oil discoveries increase GDP mainly by increasing output per worker. We also show that while both onshore and offshore discoveries increase manufacturing GDP (potentially in a mechanical way, since manufacturing includes oil production), only onshore discoveries increase services GDP and urbanization. We hypothesize that demand from well-paid oil workers is responsible for the observed increase in services and urbanization. Oil municipalities become local service and commerce hubs which benefit from improved output per worker. The treatment intensity analysis suggests that major oil discoveries have a disproportionately larger impact on the local economy.

In order to shed light on whether our results are mainly driven by local price effects or real changes in the economy, we look at recent microdata from the Brazilian employment and population censuses. We find that municipalities in which oil was discovered have larger services firms, a higher density of formal services workers, and a lower fraction of workers employed in the subsistence agricultural sector than the control group. The move from rural informal work to the formal services sector explains the observed increase in urbanization and services GDP *per capita*. We also show that wages in the services sector adjust upwards. Consequently, we find evidence for both nominal and real effects. Lastly, the density of non-oil manufacturing firms and workers is not affected by oil discoveries. Our findings, therefore, do not provide support for either the de-industrialization hypothesis of natural resource discoveries or positive agglomeration effects in the manufacturing sector, but they show that oil production has important real effects on the local economy and, in particular, on the services sector.

Our results are robust to a variety of control groups, different control variables, and a restriction of the sample period to 1940–1996. The latter is important to verify whether our results are driven by direct market effects, since from 1997 onwards royalty payments became an important part of municipal income. Lastly, we show that municipalities with oil discoveries have a higher probability of hosting major downstream oil facilities than the control group. To check whether our results are driven by these downstream facilities, we re-run the regressions excluding those municipalities which host them and find that this is not the case. This suggests that upstream production not only impact the local economy via downstream production but also has a direct effect.

Since oil is one of the world's biggest industries and it is at the center of the production network in many countries, its impact on the economy has been studied extensively. The usual approach to understanding the effects of oil relies on cross-country evidence. Several papers have shown correlations between natural resources and adverse outcomes. For instance, Sachs and Warner (1995) show that resource-exporting countries tend to have lower growth rates, while Isham, Woolcock, Pritchett, and Busby (2005) point out that resource-exporting countries have poorer governance indicators.[5] However, cross-country evidence is sensitive to changing periods, sample sizes, and covariates (for an overview of the literature, see van der Ploeg (2011)).[6] Additionally, cross-country studies usually use very aggregate variables and make it difficult to control for institutional and cultural frameworks, and for policy variation between different countries.

As a result, the literature has been shifting attention to a more detailed analysis to pin down specific mechanisms of how natural resources impact the economy. Notable papers in an emergent literature which tries to address these problems more directly are, among others, Michaels (2011), Monteiro and Ferraz (2012), Allcott and Keniston (2013), and Caselli and Michaels (2013).[7] Within-country differences in output and wages account for a substantial fraction of worldwide inequality (see, for example, Acemoglu and Dell (2010) and Moretti (2011)), and natural resources may have an important role in explaining this clustering of economic activity. The main empirical challenge, however, is to deal with

[5] Also, see Arezki and Brueckner (2011)

[6] There is also a large theoretical literature which tries to explain how natural resource abundance might affect economic and political outcomes (e.g., Krugman (1987) and Caselli and Tesei (2011)).

[7] Caselli and Michaels (2013) focus on the effects of oil windfalls on government behavior and the provision of public goods in Brazil, while Monteiro and Ferraz (2012) also use windfalls in Brazil to study local political and economic outcomes. See also Brollo, Nannicini, Perotti, and Tabellini (2013) for an analysis of fiscal windfalls in Brazil. We study the direct effects of oil discoveries instead of the indirect effect via windfalls. Also, see Acemoglu, Finkelstein, and Notowidigdo (2014) and Dube and Vargas (2013) on the local effects of resource wealth.

the endogeneity of natural resource extraction, since many unobservable factors which affect economic development might be correlated with oil production and oil discoveries (for example, see Cust and Harding (2014) for an analysis of the influence of institutions on oil exploration).[8]

Our paper stands out from the existing literature in at least two important respects: Firstly, our novel identification strategy of comparing areas with oil drilling and discoveries to those with drilling but no discoveries allows us to estimate the impact of oil discoveries on local development using a (quasi-experimental) difference-in-difference approach. Secondly, we examine the entire history of oil exploration in Brazil, while the literature limits attention mostly to post-discovery periods. Lastly, the use of worker-level data makes it possible for us to look in more detail at the exact mechanism through which oil discoveries impact local economic development.

In terms of design and results, our paper is also related to the literature on agglomeration externalities, especially the branch which investigates the impact of interventions on the concentration of economic activity (important contributions include Davis and Weinstein (2002) and Greenstone, Hornbeck, and Moretti (2010)). Similarly to our research, these papers are motivated by insights into the importance of within-country differences in output and wages. Lastly, our focus on sectoral GDP links the paper to studies on the determinants of structural transformation, particularly the ones focusing on the role of the oil sector (Kuralbayeva and Stefanski (2013) and Stefanski (2010)).

We find that oil discoveries benefits local economic development. It is important to stress, however, that we cannot comment on the aggregate impact of oil discoveries on the country as a whole. Compared to national economies, municipalities are much more open and face macroeconomic policies which are invariant to their idiosyncratic conditions. By construction, our research design rules out any effect which operates through the nominal exchange rate, for example.

[8]While institutional differences might be more pronounced at a cross-country level, they are still important at a within-country level (see Acemoglu and Dell (2010)).

This article proceeds as follows. Section 2 provides background on oil drilling and on the key institutional aspects of oil exploration in Brazil. Section 3 details the research design used to identify the impact of oil on growth. In this paper we combine several datasets which are detailed in a subsection of Section 3. Section 4 discusses the estimation strategy. Section 5 shows the results and robustness checks. Section 6 concludes.

2 Background

2.1 Oil Drilling

Oil and gas exploration is a risky business. Oil companies aim to find an oil field, which corresponds to a contiguous geographic area with oil, and they thus search for areas with specific geological characteristics to drill for oil. For instance, oil companies search for areas that contain geological structures (subsurface contortions and specific rocks) for potential trapping of hydrocarbons. Geology and related disciplines provide guidance on where to search for oil traps, and estimating the probability of discovery prior to drilling is an important aspect of petroleum exploration. However, only by drilling can the company be certain that hydrocarbon deposits really exist. Even with modern technology, the only direct way of confirming the *hypothesis* of oil presence is by drilling a well. Oil companies may invest substantially in acquiring information, only to end-up with either no discoveries or none that are profitable.

Drilled wells are classified according to the result of the attempt to find oil. A drilled well can be classified, among other categories, as a discovery well, a producer well, a dry hole, or an abandoned well (e.g., because of an accident). The likelihood of finding oil from drilling can be low, even in areas with appropriate geological characteristics, and learning-by-doing is an important aspect of the petroleum industry (Kellogg (2011)). Testing by drilling is expensive and may not reduce the uncertainty regarding the existence of oil. Numbers vary, but in a newly explored area the likelihood of successfully drilling for oil can be very low, and subjective probabilities are widely accepted in the petroleum industry

(Harbaugh, Davis, and Wendebourg (1995)). Today, an exploration well (wildcat well[9]) can have a probability as low as 10% of yielding viable oil, while a rank wildcat[10] has an even smaller chance of finding oil. Therefore, even with modern technology, drilling is not a "safe bet," since there is no guarantee that a company will find oil after drilling. Given the features of drilling, oil discovery depends both on geological characteristics and on "luck."[11] Our data support the idea that discovering oil is a kind of "lottery": For every exploration well drilled which was successful, there were many more unsuccessful ones.

A large number of factors influence drilling success, such as past drilling history, regional endowment, resource depletion, onshore vs. offshore drilling, and technological progress. While not immediately relevant to our research design, it is worth pointing out that two of those factors changed during our period of analysis: the level of technology available and the availability of conspicuous targets of hydrocarbon deposits. A more detailed discussion of oil drilling is given in Appendix B.1.

2.2 Oil in Brazil

Our period of analysis is from 1940 to 2000. During most of this period, only government-owned entities were able to explore and produce oil in Brazil. In 1938, under a dictatorship that lasted from 1937 to 1945, Federal Law n. 395/38 established state control of oil development, and not until 1997 (Federal Law n. 9,478/97) were private companies allowed to autonomously explore and produce oil in Brazil. Federal Law n. 395/38 created the CNP (in Portuguese, *Conselho Nacional do Petróleo*), the only entity responsible for exploring oil from 1938 to 1953.[12] From 1953 to 1997, only one company was allowed to

[9] A well drilled a mile or more from an area of existing oil production.

[10] A well drilled in an area where there is no existing production.

[11] According to Harbaugh, Davis, and Wendebourg (1995), "luck is obviously a major factor in exploration."

[12] According to Federal Law n. 395/38, private oil companies could operate only via concessions granted by the CNP. Anecdotal evidence suggests that it was difficult for a private oil company to operate in Brazil at that time.

drill for oil in Brazil: the government-controlled Petrobras.[13] Petrobras is an integrated exploration and production company whose activities encompass all phases of the oil supply chain. Under certain circumstances, other oil companies could explore for oil in Brazil during that period, but only in partnership with Petrobras. Following the oil crisis in 1973, Petrobras and other oil companies could sign a so-called risk contract to explore specific areas between 1975 and 1987. The terms of the contracts varied, but usual stipulated that the oil found under this type of contract could not be exported and that Petrobras could simultaneously explore an adjacent area by itself.[14] There is a sharp contrast in terms of ownership of resources between the United States and Brazil. There are thousands of oil companies, with various business models in the U.S.,[15] while oil production in Brazil has historically been linked with Petrobras's monopoly.

Local governments had little space to influence Petrobras (or CNP) on where to search for oil or the speed of drilling. For one thing, Petrobras (as a national oil company) followed national goals, which were not necessarily correlated with local-level objectives. Petrobras had a long-term goal, namely, to achieve self-sufficiency for Brazil in oil production. In addition, several factors which influence exploration activity, such as the international price of oil (e.g., Mohn and Osmundsen (2008)), are determined exogenously.[16] Furthermore, Petrobras knew it could drill only in locations with selected geological char-

[13]Petrobras was created in 1953 by Federal Law n. 2,004/53. Constitutional Amendment 09/1995 and Federal Law n. 9,478/97 changed the upstream industry in Brazil: After 1997, the upstream oil market was open to domestic and foreign oil firms, and Petrobras started to face competition. Nowadays, Petrobras is one of the largest oil companies in the world and a leading company in oil exploration, with contributions to technology, especially for deep-water exploration.

[14]The first contracts were signed in 1976 through a public bidding. Nine of the ten bidding areas were offshore, the remaining one being in the Amazon basin. More than 100 risk contracts were signed over a period of 12 years. According to the contracts, all the oil that was found was required to be sold internally until the country reached self-sufficiency in oil production. This did not occur until three decades later, in 2006.

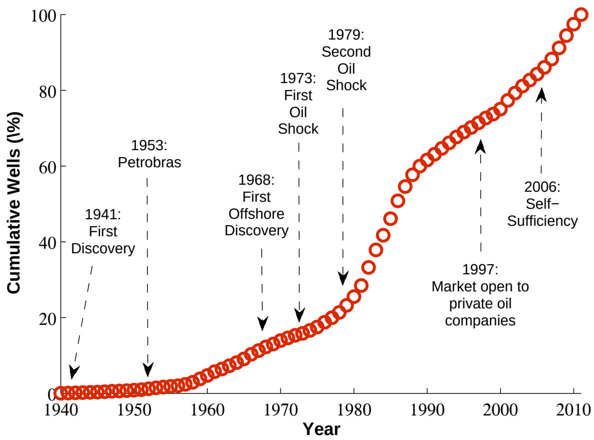
[15]Institutions such as the U.S. Energy Information Administration and the Independent Petroleum Association of America report the existence of several thousand oil operators in the U.S. economy.

[16]Figure B.1 in Appendix B.1 shows the very high correlation between drilling activity in Brazil and international oil prices.

acteristics.[17] One concern might be that Petrobras's "risk contract" partners were local companies with a local agenda. However, the vast majority of those contracts were signed with profit-maximizing multinational oil companies. Three smaller Brazilian companies also signed exploration contracts with Petrobras. One of these three companies was government owned Paulipetro, which was founded in 1979 by São Paulo state. Between 1980 and 1983, Paulipetro drilled 33 wells in one specific area; these attempts led to only one discovery well — one that was non-economical (Bosco (2003)).

The Brazilian oil sector has experienced substantial development from 1940 onwards. In 1939, the first onshore field (which was non-commercial) was discovered, and in 1941 the first onshore commercial producer well was drilled. The first oil discovery from an offshore well took place in 1968. In 2011, Brazil was the world's 13th largest producer of oil and gas, with 2.2 millon barrels per day, which represents 2.6% of the total produced worldwide. Brazil has the world's 14th largest proven petroleum reserves in the same year (ANP (2012)). The oil sector is important for the Brazilian economy: In 2011, the oil sector represented 12% of the total Gross Domestic Product (CNI (2012)). Figure 1 summarizes domestic and international events related to oil exploration and production in Brazil.

Fig. 1: Events and Oil Drilling: 1940-2011



Notes. Figure shows the cumulative percentage of oil wells drilled in Brazil during the period from 1940 to 2011.

The oil business is crucial to several municipalities. The main economic activity in

[17] Figure 4(b) shows that all oil wells in Brazil are located within sedimentary basins.

several of the ten municipalities with highest *per capita* GDP is associated with the upstream or downstream oil industry. These municipalities include São Francisco do Conde (with a refinery[18]), Triunfo (petrochemical industry) and Quissamã, Campos, and Macaé (linked to offshore production). Anecdotal evidence suggests that municipalities which discovered large amounts of oil underwent a significant transformation and substantial economic growth. For example, Macaé, a fishing municipality, was transformed from a rural area to a very urban one after Petrobras discovered offshore oil there and located some of its key production facilities in Macaé in the 1970s. There are also anecdotes of Petrobras hiring hundreds and thousands of rural workers to join drilling expeditions. In the 1960s, oil was discovered in the municipality of Carmopólis, which is located in an area which was formerly dominated by sugarcane production. Since then, Carmopólis has undergone a change in its main business due to the presence of Petrobras and related oil service companies. Carmopólis has enjoyed high GDP growth, although there have been complains regarding the lack of connections between oil service firms and the community.[19] Oil was also discovered in the municipality of Alagoinhas in Bahia in 1964. A number of successive discovery wells prompted Petrobras to locate some of its facilities in Alagoinhas in the late 1960s. Anecdotal evidence suggests that this has led to rapid economic growth in the area, particularly in the services sector. Alagoinhas has become a services hub for the surrounding municipalities and large commercial outfits located there.[20]

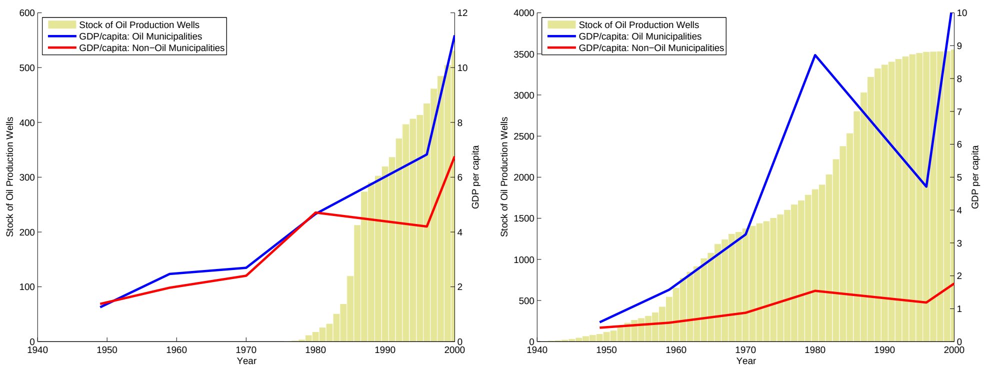
Figures 2(a) and 2(b) show GDP *per capita* for the period 1940–2000 in the states of Rio de Janeiro and Bahia (the first state in which oil was discovered), respectively. For each state, the graphs illustrate the evolution of GDP of municipalities with and without oil. It can be seen that a wedge in GDP per capita between oil-producing municipalities and those without oil production emerged over the years. Furthermore,

[18]The first refinery was constructed in 1949 in the municipality of São Francisco do Conde (located in Bahia state). The refinery, RLAM (Refinaria Landulpho Alves-Mataripe), is located near the very first wells that discovered oil in the country.

[19]See http://www.uff.br/macaeimpacto/OFICINAMACAE/

[20]See http://pt.wikipedia.org/wiki/Alagoinhas

Fig. 2: GDP *per capita* in Oil and Non-Oil Municipalities



(a) GDP *per capita* in Rio: 1940–2000 (b) GDP *per capita* in Bahia: 1940–2000

Notes. Figure shows *per capita* GDP in municipalities of the states of (a) Rio de Janeiro an (b) Bahia in which oil was discovered during the period 1940 to 2000 (blue line) and those in which it was not (red line). Rio de Janeiro is the most important producer (in terms of volume of oil), and the first oil discovery there took place in the late 1970s. The first commercial oil well in Bahia was discovered in 1941.

the timing appears to correspond quite closely to the development of the oil sector in each state. At first glance, oil production appears to have substantially increase local GDP. Two questions naturally arise from this. Firstly, is the observed correlation causal? And secondly, how did the non-oil sector develop? Since oil extraction is a high-value-added activity, local GDP increases mechanically when oil is produced, bar any extreme "Dutch Disease" effect. We are interested in assessing whether the spillovers of oil production to other sectors are positive or negative.

Only after 1997 (Federal Law n. 9,496/1997) did royalties begin to represent a significant amount of revenue to local governments. In a robustness exercise, we restrict our analysis to the years 1940–1996 to capture only the direct effect of oil production rather than the indirect effect through royalties.

In the next section, we discuss the identification strategy used to retrieve the effect of oil discoveries on growth of local economies in Brazil.

3 Research Design

We study the impact of oil by defining the analysis in terms of the treatment evaluation literature, where we see oil production as our treatment of interest and oil discoveries as the assignment to treatment. In this section, we detail our research design, which is based on exploiting the quasi-random nature of oil discoveries. Our research design exploits unconfounded assignment, and we perform several exercises to guarantee adequate overlap between the treatment and control groups (strong ignorability, as in Rosenbaum and Rubin (1983)). While it is common in the literature on natural quasi-experiments to match on observable variables, our research design additionally provides several strategies for "matching on unobservables". We start by describing the data and then discuss the exogeneity of oil discovery and its relation to the treatment assignment. We then turn to the issue of balance in the covariate distributions between treatment and control groups.

3.1 Data

The data on drilling are from *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis* (ANP), the Brazilian oil and gas industry regulator. The well dataset contains detailed information on the drilling of 20,052 wells in Brazil spanning the years from 1940 to 2000. The dataset contains the location (latitude and longitude) of each well, the exact date of the drilling, and the result (whether oil was found, whether the well is a dry hole, whether only water was found, or whether the well was abandoned because of an accident.)[21] Furthermore, we have information on the viability of exploring the oil deposit (when oil was found), and on whether the oil company started production.

The richness of the well dataset allows us to study several possibilities regarding the stages of oil extraction and production (the upstream oil industry). From the data, we are able to separate places where drilling took place ($J = 1$) from places with no drilling ($J = 0$). We are also able to obtain information on places with oil discoveries (Z) and with oil production (D). As a first step, we created a dummy variable for drilling (J), two different dummy variables for discovery (Z), and a dummy for well production (D).

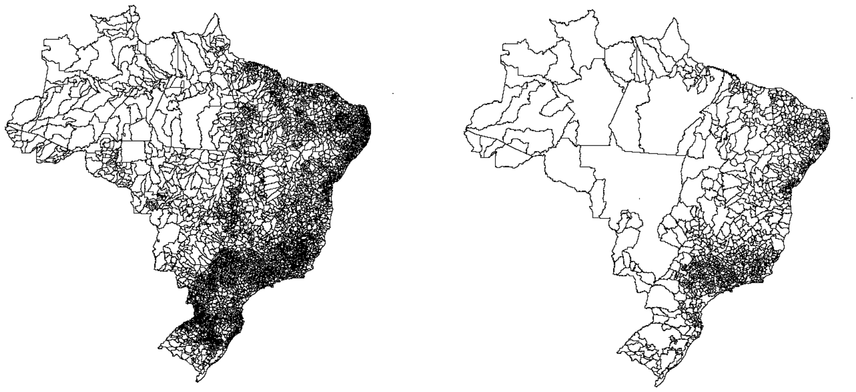
[21] We obtained more the 50 different classifications from the dataset, but we were able to aggregate all of them into a few major categories (see Table 2). The data differentiate between oil well, gas well, and oil and gas well. One limitation of the dataset is that it does not include information on the amount of oil produced by each individual producer well during the period of interest. Data on well production are available only from the the 2000 onward.

The dummies for drilling and production follow immediately from the well data. The drilling dummy is set equal to one when at least one well was drilled in the municipality, and the production dummy is set equal to one when there is at least one producer well in the municipality. In terms of discoveries, there are several possibilities, as the data allow us to differentiate between a field discovery, a subfield (reservoir) discovery, and a field extension discovery. We define two different discovery dummies as follows. The first dummy ("All Discoveries") is set equal to one when at least one field, subfield, or field extension discovery was made in the municipality. The second dummy ("True Discoveries") is set equal to one when at least one field or subfield discovery and at least one field extension discovery were made in the municipality. The rationale for the latter is that any substantial discovery includes a field or subfield discovery and subsequent field extension discoveries to delineate the size of the oil field (see Appendix B.1). For now, we will use the "All Discoveries" dummy to start with the most general possible definition of discoveries.

The Brazilian federation has three administrative levels: federal government, states, and municipalities. One complication when dealing with municipalities in Brazil is the process of detachments and splits that have taken place over the years. For instance, in 1940 there were 1,574 municipalities, while in 1997 there were 5,507. In order to deal with the detachments, we used the concept of a Minimum Comparable Area (MCA), which consist of sets of municipalities whose borders were constant over the study period. Our data were aggregated to 1,275 MCAs in 1940. Figure 3 shows the boundaries of municipalities in 1997 and the corresponding MCAs in 1940. Additional information on MCA aggregation can be found in Da Mata, Deichmann, Henderson, Lall, and Wang (2007).

We allocate the wells to the MCAs as follows. For onshore wells, we simply allocate the wells to the MCAs within whose boundaries they were located. For offshore wells, we calculate the distance from each well to the nearest coastal MCA and allocate the offshore well to that MCA. As a robustness check, we also use an alternative method to allocate offshore wells to MCAs (see Subsection 5.2).

Fig. 3: Municipalities and Minimum Comparable Areas (MCAs)



(a) 5,507 Municipalities in 1997 (b) 1,257 MCAs in 1940

Notes. Figure 3(a) shows the administrative boundaries of the 5,507 municipalities that existed in 1997 in Brazil. Figure 3(b) shows the aggregation to the 1,275 Minimum Comparable Areas (MCAs) in 1940.

Table 1: Number of Discoveries by Decade

	# of Wells: Discoveries			Units Assigned to Treatment		
Decade	Total	Onshore	Offshore	Total	Onshore	Offshore
1940	9	9	0	3	3	0
1950	48	48	0	8	8	0
1960	212	206	6	19	18	1
1970	203	117	86	13	4	9
1980	671	434	237	15	7	8
1990	285	158	127	6	2	4

Notes. Data from ANP (Brazilian oil and gas industry regulator). The units assigned to treatment are Minimum Comparable Areas (MCAs). MCAs consist of sets of municipalities whose borders were constant over the study period.

Table 1 shows the number of wells discovered by decade. It contains information on the total number of discoveries, and on onshore and offshore discoveries. It also has information on the total number of units assigned to treatment over time. Table 2 shows the number of wells by category. Wells are classified broadly as exploratory wells and development wells. Exploratory wells are drilled to test for the presence of oil, while wells drilled inside the known extent of the field are called development wells (e.g., producer wells).[22] Unsuccessful drilling is classified as a dry hole in both exploratory and development categories. See Appendix B.1 for a detailed explanation of the types of wells.

We have the following numbers regarding oil discoveries in Brazil:

- Total number of MCA units = 1,275

[22]Note that the two instruments (true discoveries and all discoveries) apply to exploratory wells.

Table 2: Number of Wells by Category

Classification	Category of Well	Offshore	Onshore	Total
Exploratory Wells	Discovery of New Field	129	304	433
	Discovery of New Subfield (Reservoir)	88	234	322
	Discovery of Field Extension (Step-out)	258	419	677
	Dry Hole	1,067	2,556	3,623
Development Wells	Producer	1,368	9,101	10,469
	Carries Oil or Gas	7	1	8
	Production Not Feasible	327	521	848
	Injection of Water, Steam, or Gas	201	774	975
	Dry Hole	73	1,017	1,090
Other	Abandoned	421	554	975
	Special	62	369	431
	Missing Category	30	171	201
	Total	4,031	16,021	20,052

Notes. Data from ANP (Brazilian oil and gas industry regulator). Wells are classified broadly as exploratory wells and development wells. Exploratory wells are drilled to test for the presence of oil. If the exploratory drilling was proven unsuccessful, the well is classified as a dry hole. Wells to delineate the extension of the oil field (step-out wells) are also classified as exploratory wells. Every well drilled inside the known extent of the field is called a development well (e.g., producer wells and injection wells). In the development well category, unsuccessful drilling is also classified as a dry hole. Special wells are water wells or the ones used for mineral research and experiments.

- All Discoveries MCAs = 64
- True Discoveries MCAs = 45
- Dry hole MCAs = 158
- Neighbors of discovery MCAs = 156

We work with three main outcome variables: population density, the urbanization rate,[23] and *per capita* GDP (overall as well as sectoral). Data on total population, population located in urban areas, total area of the municipality, as well as data on employment (total and sectoral) come from historical population censuses. Data on municipal gross domestic product (GDP) and on the shares of manufacturing, agriculture, and services in GDP are from Ipeadata.[24] Using this information, we construct our outcome variables to obtain a panel from 1940 to 2000. In 1941, the first well started to produce oil, so 1940 is our pre-treatment year. The panel data are balanced, and we do not observe any attrition. However, the time dimension is unequally spaced for GDP *per capita*. Because Population Censuses where historically only conducted every 10 years and there are no data on GDP for 1990 or 1991, we end up with GDP *per capita*

[23]The urbanization rate is the proportion of the population living in urban areas.

[24]The GDP calculations are detailed in Reis, Tafner, Pimentel, Serra, Reiff, Magalhaes, and Medina (2004). The GDP is deflated using the national implicit price deflator. In subsection 5.1, we use the composition of GDP to argue that we capture a variation in real local GDP instead of a price effect by showing that oil municipalities undergo an important structural transformation.

data for the years 1949, 1959, 1970, 1980, 1996, and 2000. By contrast, our panel is virtually equally spaced in time for the other two dependent variables (urbanization rate and population density): 1940, 1950, 1960, 1970, 1980, 1991, 1996, and 2000.

Additionally, we collected data on average temperature, average rainfall, and average altitude from Ipeadata.[25] Further data comprise the latitude and longitude of each MCA as well as geographical indicators of its location (on the coast, in the Amazon region, and in the semiarid region).[26] Table A.3 in Appendix A shows the summary statistics of the variables used in the analysis.

In further analysis we use microdata from the employment and population censuses. The Brazilian Ministry of Labor's RAIS (Relação Anual de Informações Sociais) provides matched employer–employee microdata (see data description in Appendix D).

3.2 Treatment Assignment

Municipalities in which oil was discovered are assigned to treatment. The untreated (control) group comprises the locations with drilling but no oil discoveries. Our treatment assignment process is very similar to a randomization: Several attempts to drill oil were made, but nature has endowed only some places with oil. Drilling took place in locations with selected geological characteristics, with little room for influence by local governments. Figure 4(b) shows that oil drilling in Brazil is concentrated in sedimentary basins. Since the locations of oil reserves are determined by geology, selection into treatment is unlikely or impossible. In other words, municipalities had no control over the assignment mechanism and thus could not influence their treatment regime.

Note that there is some noncompliance with the assigned treatment, that is, in some locations oil discovered ($Z = 1$) but no oil was produced ($D = 0$). We have information on whether a discovered oil field is economically viable to begin production. Viability

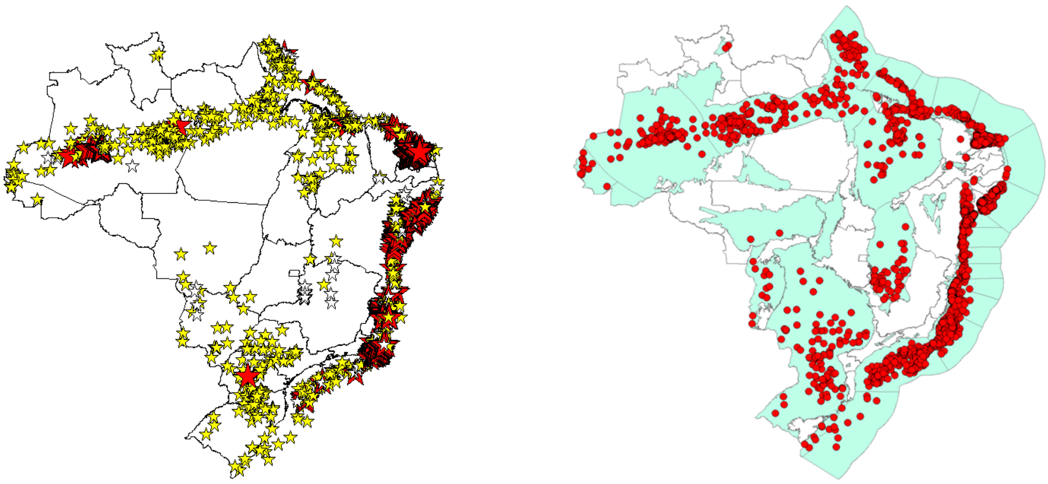
[25] Temperature is measured in degrees Celsius, precipitation in millimeters per month, and altitude in meters.

[26] To construct the shapefile of 1940 MCAs, we combined the shapefile of 1997 municipalities with the matching between the 1940 MCAs and the corresponding 1997 municipalities. From the shapefile of 1940 MCAs, we constructed latitudes, longitudes, and geographical indicators.

depends to the largest extent on the characteristics of the oil field but potentially also on some local characteristics. Part of the costs of producing oil may be systematically correlated with unobservable local characteristics. For instance, existing infrastructure and institutional support from the local and state governments might influence the decision to produce oil at the margin. As a result, the research design implies random assignment of locations to treatment and control groups, but allows for non-random selection of participants into treatment (once assigned to treatment). As part of our empirical strategy, we thus use discoveries as an instrumental variable for production.[27]

For our identification strategy to be valid, we need to show that (the intensity of) drilling attempts are exogenous to local characteristics (conditional on appropriate geographical controls) and that (conditional on drilling taking place) the discovery of oil is a "lottery."

Fig. 4: Oil Wells in Brazil: 1940-2000



(a) Oil Wells: Discovery (red) vs. dry (yellow)   (b) Sedimentary Basins (light green)

Notes. The figures show the locations of approximately 20,000 drilled wells (the universe of wells drilled in Brazil during the period from 1940 to 2000). In Figure 4(a), wells with Oil Discovery are in red, Dry Wells are in yellow, and others are in white. Figure 4(b) shows the locations of sedimentary basins in Brazil (in light green). Both figures show the administrative boundaries of the 27 states of Brazil that have been in effect since 1988. (See https://www.youtube.com/watch?v=_ZKdnUeBcOI for a short video on the geographic distribution of drilling activity in Brazil from 1940 to 2000.)

[27]Part of the non-compliance is due to MCAs discovering oil towards the end of our sample period but not starting production until after 2000.

3.3 Assessing the Design

Our research design is based on the idea that drilling took place only in locations with selected geological features, with no influence from local governments. Thus far we have discussed several points that support the exogenous nature (from the viewpoint of local economies) of drilling in Brazil: the risky characteristics of oil exploration, the self-sufficiency goal of Petrobras, and the concentration of drilling attempts in geological target areas in the Amazon and on the coast (see Figure 4). We now provide further evidence of a lack of a relationship between drilling and local characteristics.

Table 3: Correlation between Drilling Attempts and Pre-Treatment Characteristics

	(1)	(2)	(3)	(4)
Dependent variable:	**Drilling Dummy**		**Drilling Count**	
	Linear Probability	Logit	Linear Probability	Poisson
Urbanization in 1940	0.0575	0.481	28.32	1.284
	(0.0939)	(0.837)	(22.22)	(0.888)
Pop. Density in 1940	-0.000343	-0.00171	-2.722	-0.177
	(0.000249)	(0.00161)	(3.354)	(0.167)
GDP *per capita* in 1949	-0.00712	-0.0787	3.413	0.129
	(0.0144)	(0.156)	(8.567)	(0.404)
Semiarid Indicator	0.00742	0.0938	20.63	1.292*
	(0.0220)	(0.232)	(19.95)	(0.782)
Amazon Indicator	0.395***	2.292***	-7.137	-0.809*
	(0.0530)	(0.276)	(7.567)	(0.470)
Coastal Indicator	0.518***	2.776***	90.65***	3.001***
	(0.0443)	(0.243)	(34.54)	(0.651)
Constant	0.0934***	-2.314***	3.725	1.572***
	(0.018)	(0.184)	(8.538)	(0.374)
Observations	1,275	1,275	1,273	1,273
R-squared	0.255	-	0.053	-

Notes. Robust standard errors in parentheses. The regressions are for 1,275 Minimum Comparable Areas (MCAs). There are two dependent variables: a dummy variable that is equal to 1 if any drilling attempt was made during the years 1940 to 2000 (columns (1) and (2) of the table) and the number of drilling attempts made during that period (columns (3) and (4) of the table). The pre-treatment variables are urbanization rate in 1940, population density in 1940, and *per capita* GDP in 1949. The geographical controls are indicator variables showing whether the MCA is located in the semiarid region, in the Amazon region, or on the coast.
*** p<0.01, ** p<0.05, * p<0.1

Table 3 shows simple regressions between drilling attempts and pre-treatment characteristics. We consider our three main outcome variables (population density, urbanization, and *per capita* GDP) in the 1940s. We construct two variables related to drilling: a dummy that is set equal to 1 for each MCA in which at least one drilling attempt was made in 1940–2000 and another that equals the number of drilling attempts made in each MCA. Using different models, we show that drilling attempts are uncorrelated with initial economic conditions. The correlations given in Table 3 strongly support the patterns from Figure 4: drilling is determined by geological and geographic characteristics, and not by pre-treatment population, GDP, or urbanization.

Table 4: Discovery Dummies: Analysis

Dependent Variable:	Oil Production Dummy (1)	Oil Production Dummy (2)
All Discoveries Dummy	0.681***	
	(0.0524)	
True Discoveries Dummy		0.777***
		(0.0472)
MCA FE	Yes	Yes
Year FE	Yes	Yes
Observations	8,901	8,901
Number of MCAs	1,273	1,273
Geographical Controls	Yes	Yes
Initial Conditions	Yes	Yes
Estimation	FE	FE
F-Statistics	9.86	20.41

Notes. Standard errors clustered at the MCA level. The explanatory variables are the two dummies related to oil discovery. The geographical controls and initial conditions have time-varying coefficients. The initial conditions with time-varying coefficients are *per capita* GDP in 1949, urbanization rate in 1940, and population density in 1940. The geographical controls with time-varying coefficients are latitude and longitude, the Amazon dummy, and the coastal dummy. Total sample: 1,275 Minimum Comparable Areas (MCA). *** p<0.01, ** p<0.05, * p<0.1

As mentioned earlier, we devise two different ways to capture discoveries. Table 4 compares the predictive power of the "All Discoveries" and "True Discoveries" dummies for explaining production. We include MCA and year fixed-effects as well as the initial economic conditions and baseline geographical controls with time-varying coefficients. The "True Discovery" dummy is more closely related to production. It has the higher t-statistic and F-statistic, and its coefficient also turns out to be larger. Since any substantial field discovery will be followed by a field extension discovery, it is not surprising that the "True Discovery" dummy is more closely related to actual production.

For the "True Discovery" dummy to be valid, it is not sufficient to show that drilling is uncorrelated with initial conditions; we also have to check whether (conditional on a discovery) additional drilling is also unrelated to local economic development. Specifically, if Petrobras, following an initial discovery, tried harder to find a field extension discovery in a location which was growing fast, or which had high demand, this would bias our results. Table 5 shows that this is not the case. Unsurprisingly, drilling attempts increase significantly after an initial discovery was made in an MCA. A first discovery is a strong signal, and naturally Petrobras subsequently intensifies its efforts in that particular area. Importantly, however, there is no indication that drilling increases *more* in MCAs with higher GDP *per capita*, more urbanized MCAs, or more densely populated ones. Both initial drilling attempts and follow-up drilling are thus orthogonal to local economic conditions.

Table 5: Drilling, Conditional on a Field Discovery

	(1)	(2)
Dependent variable:	Wells drilled per year	
Estimation:	OLS	Poisson
Field Discovery Dummy	5.502**	5.255***
	(2.259)	(0.514)
Field Discovery Dummy * log Population Density	-0.517	-0.0689
	(0.600)	(0.0721)
Field Discovery Dummy * log GDP/capita	0.849	0.107
	(1.121)	(0.135)
Field Discovery Dummy * Urbanization	4.706	0.690
	(5.925)	(0.829)
Constant	0.0285***	-3.557***
	(0.0104)	(0.366)
Observations	5,098	5,098

Notes. Robust standard errors in parentheses. The regressions are for 1,275 Minimum Comparable Areas (MCAs). The dependent variable is the count of drills per year. The explanatory variables are a dummy for a field discovery and the interactions between this dummy and GDP/capita, urbanization, and population density.
*** p<0.01, ** p<0.05, * p<0.1

Until this point, we have been concerned with the exogeneity of drilling attempts to local economic conditions. For our identification strategy to be valid, we also have to show that (conditional on drilling) discoveries are unrelated to local economic characteristics. We restrict the sample to only those municipalities which drilled for oil, and we find that both the number of discoveries and the ratio of successful drilling to unsuccessful drilling are unrelated to local economic characteristics. Table 6 shows thus that (conditional on drilling taking place) pre-treatment economic characteristics do not influence drilling success. It is in fact particularly reassuring that the success ratio is uncorrelated with all controls, that is, conditional on drilling taking place, success is truly a lottery.[28]

[28] We repeat the analysis of Tables 3, 5, and 6 by running separate regressions for each of the three pre-treatment characteristics. The regressions show that (conditional on geography) each pre-treatment characteristic individually is unrelated to the dependent variables.

Table 6: Discoveries, Conditional on Drilling

	(1)	(2)	(3)
Dependent variable:	**Number of Discovery Wells**		**Drilling Success Ratio**
	Linear Probability	Poisson	Linear Probability
Urbanization in 1940	0.524	-0.844	0.121
	(9.766)	(1.478)	(0.125)
Pop. Density in 1940	0.435	0.108	-0.00255
	(0.912)	(0.169)	(0.0165)
GDP *per capita* in 1949	2.779	0.548	-0.00148
	(2.302)	(0.381)	(0.0300)
Semiarid Indicator	10.53	1.362**	0.104
	(7.595)	(0.562)	(0.0679)
Amazon Indicator	2.499	-0.377	-0.0263
	(3.733)	(0.746)	(0.0586)
Coastal Indicator	10.77**	1.704***	0.0595
	(5.190)	(0.535)	(0.0390)
Constant	0.783	0.834*	0.0622
	(3.059)	(0.471)	(0.052)
Observations	222	222	210
R-squared	0.070	-	0.031

Notes. Robust standard errors in parentheses. The regressions are for the 222 Minimum Comparable Areas (MCAs) in which Petrobras drilled for oil. The drilling success ratio is the ratio of exploratory wells with oil to exploratory dry wells. The pre-treatment variables are urbanization rate in 1940, population density in 1940, and *per capita* GDP in 1949. The geographical controls are indicator variables showing whether the MCA is located in the semiarid region, in the Amazon region, or on the coast.
*** p<0.01, ** p<0.05, * p<0.1

3.4 Assessing the Overlap of Covariates

Our baseline strategy to control for unobservables is to use municipalities where there was drilling for oil but no discovery as our control group. However, even if a place where oil was discovered is sort of a "lottery winner", which would guarantee unconfoundedness, a lack of overlap (or common support) could still be a threat to internal validity. Figure 4 shows that oil deposits are not randomly distributed across the country, but rather concentrated in the basin of the Amazon River (onshore wells) and on the Atlantic coast (offshore wells).

We investigate systematic differences between the group assigned to treatment and the control group. Rubin (2001) proposes a set of criteria to check for overlap. In this paper, we use the normalized (or standardized) difference to assess the difference in location in the covariate distributions (Imbens and Wooldridge (2009)). Standardized differences are not influenced by sample size, unlike t-tests and other statistical tests.

We present detailed results of this assessment in Appendix A. Our dry-drilling control group presents a good performance in terms of pre-treatment characteristics such as urbanization and population density. However, it does not pass the standardized difference assessment for some geographical controls such as longitude and coastal dummy. In fact, the dry-drilling group is more spread out over the Brazilian territory. As a result, to improve overlap, we created a matched subsample of the "drilling but no discovery" group. Propensity score matching (or trimming) is a common way to improve overlap (Imbens and Wooldridge (2009)). For this subsample, we choose the 64 municipalities with the highest propensity score and call this control group "matched dry drilling."[29] It should be emphasized that while using this matching may improve internal validity, it may reduce the external validity of the results, because we are now focusing on a subset of the original sample (Imbens and Wooldridge (2009)).

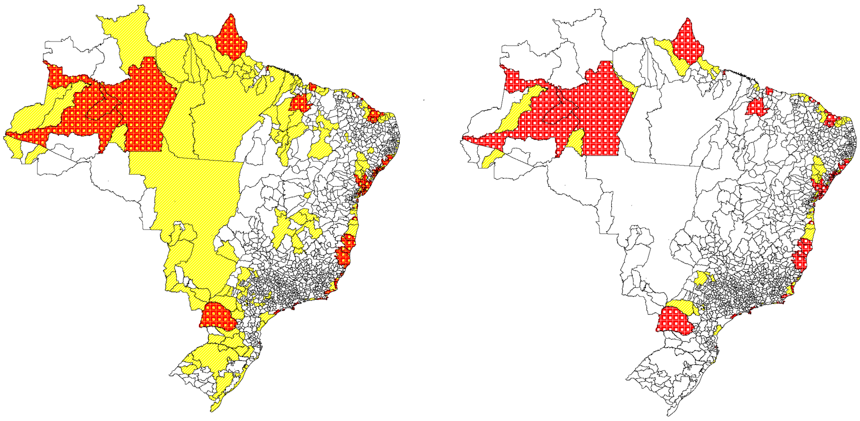
As an alternative, we also use direct neighbors as one of our control groups. This is a strategy widely employed in the literature. Neighbors are likely to have similar geographical and institutional characteristics and are likely to be very similar across other unobservables. Additionally, we consider all non-oil MCAs in oil states, all dry-drilling MCAs which are not neighbors of discovery MCAs (dry drilling, no neighbor), and a trimmed subsample of the neighboring MCAs. The idea is to create multiple comparison groups to strengthen the results.

Figure 5 shows maps with the locations of the two most relevant control groups. Figure 5(a) shows the places with discoveries and the set of MCAs where drilling took place and no oil was found. Figure 5(b) displays the matched dry-hole subpopulation.

An implicit assumption in the analysis is the stable unit treatment value assumption, that is, that there is no spillover effects of the treatment on the control group. In the presence of spillover effects, neighboring locations may also receive part of the treatment. To alleviate doubts about spillovers, we have included the "dry drilling, no neighbor"

[29]The pre-treatment characteristics used in the propensity score model include: population density in 1940, urbanization rate in 1940, GDP *per capita* in 1949, share of manufacturing out of the total GDP in 1949, share of services in 1949, share of agriculture in 1949, geographical indicators (whether the MCA is located on the coast, in the semiarid region, or in the Amazon region), historical average rainfall and temperature, and latitude and longitude. Since the very large share of relevant discoveries took place after the creation of Petrobras in 1953 (recall from Table 1 that oil was discovered in only 9 wells during the 1940s), the GDP variable for 1949 can be considered pre-treatment.

Fig. 5: Treatment and Control Groups



(a) Discovery (Red), Drilling (Yellow) (b) Discovery (Red) and Matched Dry Drilling Sample (Yellow)

Notes. Figures show 1,275 Minimum Comparable Areas (MCAs) in 1940. The discovery dummy is the "All Discoveries" dummy (which is equal to one when at least one field, subfield, or field extension discovery was made in the MCA).

group as one of our control groups. The next section discusses the empirical strategy used to recover the main estimand of interest.

4 Estimation

We now briefly discuss the empirical strategy we use to recover the impact of oil discoveries. The estimand of interest is the Intention-to-Treat (ITT): the average impact of being assigned to treatment. Let y_i be the potential outcome for local economy i, and let the indicator of treatment assignment be $Z_i = \{0, 1\}$. The ITT estimand is represented by $\text{ITT} = \mathbb{E}[y_i|Z_i = 1] - \mathbb{E}[y_i|Z_i = 0]$.

In the discussion below, the oil discovery dummy is represented by Z_{it} (treatment assignment), which is set equal to 1 if oil was discovered in MCA unit i in period $t \geq \bar{t}$, where $\bar{t}$ is the time of the discovery. A regression using Z_{it} is an intent-to-treat (ITT) analysis. We assume an additive and linear empirical specification to estimate an ITT effect, as follows:

$$Y_{it} = \alpha + \tau_{_{ITT}} Z_{it} + \beta_t' X_i + \gamma_i + \rho_t + \epsilon_{it}, \tag{1}$$

where Y_{it} is the outcome variable, X_i are time-invariant MCA characteristics, including the pre-treatment level of the dependent variables, ϵ_{it} is an error term, ρ_t denotes year fixed effects and γ_i denotes MCA fixed effects. The time t ranges from 1940 to 2000. The (exogenous) source of cross-sectional and time variation is given by the discovery of oil in unit i at time t. As a result, the parameter τ_{ITT} should capture an intent-to-treat effect. Note that ITT is a lower bound on the average treatment effect. We add γ_i to capture time-invariant characteristics and ρ_t to capture common aggregate shocks that hit all locations.

We also use a set of additional covariates X_i in equation (1). Recall that we trim by using the propensity score to create some control groups for robustness checks. After matching by using the propensity score, model dependence is not eliminated but will normally be reduced. Parametric procedures have the potential to improve causal inferences, even after matching when the match is not exact (Ho, Imai, King, and Stuart (2007)). Moreover, the trimming used to create the control groups also helps with the common trend assumption. Lastly, note that policy variation takes place at the MCA level, and errors within the spatial units may be correlated. Therefore, standard errors are clustered at the MCA level in all regressions (Bertrand, Duflo, and Mullainathan (2004)).[30]

5 Results

This section is divided into three parts. The first part discusses the baseline results and a host of robustness checks regarding the effects of oil discoveries. In the second part, we provide additional results which compare onshore to offshore discoveries, retrieve the local average treatment effect of oil production, and explore the link between upstream and downstream oil production. The third part uses microdata from the 2000 employment and population censuses to explore the mechanism underlying our results in more detail.

[30] Time could threaten identification if discoveries took place in boom periods: Places where oil was discovered during a boom may have had a better opportunity to promote local growth. Our use of year fixed effects helps to alleviate this concern. Additionally, the bulk of drilling activity (and some important discoveries) took place in the 1980s, a decade labeled as the "lost decade" because of its low GDP growth. In other words, important discoveries did not take place during boom periods in Brazil.

5.1 Baseline Results

As discussed in the estimation section (see Section 4), we include MCA as well as year fixed effects, and clustered standard errors at the MCA level, in all regressions. Additionally, we control for geographical characteristics and initial conditions with time-varying coefficients. The controls included in all regressions are: *per capita* GDP in 1949, urbanization rate in 1940, population density in 1940, latitude and longitude, the Amazon dummy, and the coastal dummy.

Table 7: Intention-to-Treat Effect of All Oil Discoveries: Socio-Economic Outcomes

	Dry Drilling			Matched Dry Drilling		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Population Density	ln GDP *per capita*	Urbanization Rate
Discovery Dummy	-0.0390	0.125*	0.0283	-0.0400	0.146*	0.0253
	(0.0579)	(0.0728)	(0.0187)	(0.0626)	(0.0783)	(0.0199)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,776	1,332	1,776	1,024	768	1,024
Number of MCAs	222	222	222	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. *** p<0.01, ** p<0.05, * p<0.1

Results for Socio-Economic Variables. Table 7 shows the baseline ITT results using the "All Discovery" dummy as our treatment assignment. We show results for both our preferred control group (dry drilling) and the matched dry drilling sample. The key independent variable is a dummy, and both *per capita* GDP and population density are expressed as logs. Therefore, the coefficient in those regressions can be interpreted as a percentage change. Urbanization is a rate bounded between 0 and 1, so that the coefficient for oil discoveries can be interpreted as a change in percentage points. GDP *per capita* increased by 12.5–14.6% over a 60-year period as a result of oil discoveries. Population density and the urbanization rate are unaffected by oil discoveries in this specification.

As discussed earlier, the "All Discovery" dummy has some drawbacks, both conceptually and in terms of its ability to predict oil production. The "True Discoveries" dummy excludes both MCAs where oil was discovered but there were no follow-up discoveries (i.e., the oil field was very small) and MCAs where there was no field discovery but only a field extension (i.e., the bulk of the field lies in a different municipality).[31]

Table 8: Intention-to-Treat Effect of True Oil Discoveries: Socio-Economic Outcomes

	Dry Drilling			Matched Dry Drilling		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Population Density	ln GDP *per capita*	Urbanization Rate
Discovery Dummy	-0.00864	0.246***	0.0443**	-0.0127	0.259***	0.0430**
	(0.0676)	(0.0856)	(0.0202)	(0.0731)	(0.0910)	(0.0213)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,776	1,332	1,776	1,024	768	1,024
Number of MCAs	222	222	222	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients.
*** p<0.01, ** p<0.05, * p<0.1

Table 8 shows the baseline ITT results using our preferred treatment assignment ("True Discoveries"). Unsurprisingly, the coefficients are markedly higher than in Table 7. The increase in *per capita* GDP is estimated at 24.6–25.9%. While population density is not significantly affected, urbanization increases by 4.3–4.4% points over the period as a consequence of oil discoveries. In other words, when we compare municipalities with significant discoveries to municipalities where Petrobras drilled for oil and either did not find any or made no substantial discovery, we find a strong positive impact on *per capita* GDP and urbanization.

Robustness. Firstly, we verify that changing the time period to 1940–1996 does not change the results. Table 9 shows that the results are virtually the same when we set 1996 as the final year. This is important, because it supports the claim that our findings are driven by the direct effect of oil production rather than the indirect effect through royalties (recall the discussion in Subsection 2.2).

The results are also both quantitatively and qualitatively robust to using alternative

[31] Implicitly, other recent papers on the impacts of oil abundance have also defined relevant discoveries. For example, Michaels (2011) uses a threshold of 100 millions barrels of reserves, and Allcott and Keniston (2013) use a cutoff in production of US$100 per inhabitant.

Table 9: Intention-to-Treat Effect of Oil Discoveries: Robustness with 1996 as final year of analysis

	Dry Drilling			Matched Dry Drilling		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Population Density	ln GDP *per capita*	Urbanization Rate
Discovery Dummy	-0.0291	0.200**	0.0459**	-0.0242	0.225**	0.0449**
	(0.0645)	(0.0926)	(0.0203)	(0.0698)	(0.0969)	(0.0210)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,776	1,332	1,776	1,024	768	1,024
Number of MCAs	222	222	222	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. The number of observations is smaller because the final year in the panel is 1996 instead of 2000. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as "True Discovery".
*** p<0.01, ** p<0.05, * p<0.1

control groups (see Table 10). Our additional control groups are all non-oil MCAs in oil discovery states, dry-drilling MCAs which are not adjacent to discovery MCAs (which we call dry drilling, no neighbor), all MCAs which are adjacent to discovery MCAs, and a matched subsample of adjacent MCAs (matched neighbors). The results for the dry drilling, no neighbor control group are reassuring in the sense that any potential spillovers should be particularly limited for this group. The matched neighbors group, on the other hand, is susceptible to spillovers but offers a good control group in terms of observable MCA characteristics. Overall, the results are remarkably similar across control groups, perhaps highlighting that our controls and the parametric fitting (the linear and additive specification represented by Equation (1)) are doing a good job in providing a precise estimate of the effects of oil on the municipalities in Brazil.[32] The estimate for *per capita* GDP ranges from 19.5–27.7% while urbanization is estimated to increase 3.6–5.2% as a consequence of oil discoveries.[33]

Table 11 shows that our baseline results are also robust to including the additional geographical controls which are available, namely, average temperature and average rainfall over the last 50 years, average altitude of the MCA, and a dummy for being located in a semiarid region. The impact of oil discoveries on *per capita* GDP is marginally lower than

[32] The results are also robust to the exclusion of major urban centers, i.e., state capitals.

[33] We also constructed trimmed (rather than matched) subsamples of the dry-drilling and neighbors control groups. The results are robust to using those.

Table 10: Intention-to-Treat Effect of Oil Discoveries: Robustness to Alternative Control Groups

	Non-Oil Municipalities in Oil States			Dry Drilling, No Neighbors		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Population Density	ln GDP *per capita*	Urbanization Rate
Discovery Dummy	-0.0560	0.262***	0.0519***	-0.0302	0.195**	0.0362*
	(0.0610)	(0.0781)	(0.0190)	(0.0751)	(0.0906)	(0.0214)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,200	4,649	6,200	1,344	1,008	1,344
Number of MCAs	775	775	775	168	168	168
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

	All Neighbors			Matched Neighbors		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Population Density	ln GDP *per capita*	Urbanization Rate
Discovery Dummy	0.0114	0.247***	0.0434**	0.0341	0.277***	0.0419**
	(0.0641)	(0.0819)	(0.0195)	(0.0645)	(0.0863)	(0.0206)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,760	1,320	1,760	1,024	768	1,024
Number of MCAs	220	220	220	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as "True Discovery".
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Intention-to-Treat Effect of Oil Discoveries: Robustness to Adding more Geographic Controls

	Dry Drilling			Matched Dry Drilling		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Population Density	ln GDP *per capita*	Urbanization Rate
Discovery Dummy	-0.00147	0.218**	0.0372*	-0.0165	0.217**	0.0390*
	(0.0723)	(0.0885)	(0.0216)	(0.0808)	(0.0944)	(0.0231)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,776	1,332	1,776	1,024	768	1,024
Number of MCAs	222	222	222	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Additional geographical controls are average temperature, average rainfall, average altitude and the semiarid dummy. Discovery is defined as "True Discovery".
*** p<0.01, ** p<0.05, * p<0.1

in the analogous regressions without the additional controls. However, since the overall fit barely improves and the coefficients for the additional controls tend to be insignificant, we prefer to exclude them to avoid a problem of over-controlling. Either way, including them only somewhat changes the results quantitatively but not qualitatively.

Sectoral GDP Results. While the results for urbanization point in a different direction, there might be a concern that the increase in GDP *per capita* is purely mechanical, in the sense that there are no spillovers from oil production to other sectors of the economy. To investigate this, Table 12 shows the impact of oil discoveries on sectoral GDP. GDP is broken up into manufacturing, services, and agriculture. Natural resource extraction is included in the manufacturing sector. While ideally we would like to decompose this further, the available data does not allow us to do so. As such, it is not surprising or particularly insightful that manufacturing GDP increases significantly with oil discoveries. Importantly, however, services GDP increases by about 20%, while agricultural GDP is unaffected. These results are interesting for two reasons. First of all, it is reassuring (in terms of our research design) that agricultural GDP is not affected. An increase in agricultural GDP might have raised doubts that we are mainly picking up local price effects rather than changes in real municipal GDP. Secondly, the results suggest that there are spillovers from oil discoveries to the services sector. One candidate for a channel is direct demand from oil firms and high-paid oil workers.

Table 12: Intention-to-Treat Effect of Oil Discoveries: Sectoral GDP *per capita*

	Dry Drilling			Matched Dry Drilling		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Manufacturing GDP *per cap*	Services GDP *per cap*	Agriculture GDP *per cap*	Manufacturing GDP *per cap*	Services GDP *per cap*	Agriculture GDP *per cap*
Discovery Dummy	0.449**	0.213**	0.0569	0.456**	0.215**	0.0664
	(0.182)	(0.0968)	(0.107)	(0.189)	(0.104)	(0.109)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,325	1,321	1,328	765	764	765
Number of MCAs	222	222	222	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as “True Discovery”.
*** p<0.01, ** p<0.05, * p<0.1

Output per Worker. To investigate the sectoral GDP results in more detail, we used historical censuses to collect data on sectoral employment by municipality going back to 1940. We then constructed a measure of output per worker by dividing sectoral GDP by sectoral employment for every MCA.[34],[35]

Table 13: Intention-to-Treat Effect of Oil Discoveries: Sectoral Output per Worker

	Dry Drilling			Matched Dry Drilling		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Manufacturing Y/L	Services Y/L	Agriculture Y/L	Manufacturing Y/L	Services Y/L	Agriculture Y/L
Discovery Dummy	0.265*	0.221**	-0.0717	0.222	0.188*	-0.0535
	(0.139)	(0.106)	(0.0881)	(0.143)	(0.113)	(0.0871)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,533	1,542	1,547	883	891	891
Number of MCAs	222	222	222	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as "True Discovery".
*** p<0.01, ** p<0.05, * p<0.1

Table 13 shows that oil discoveries increase output per worker in the manufacturing sector by slightly over 20% (recall again that this includes oil production) and by roughly 20% in the services sector. The agricultural sector in not affected. While the result is significant for the services sector for both control groups, it is marginally insignificant at conventional levels in one of the two regressions for the manufacturing sector. Comparison of the estimated coefficients with the increases in sectoral GDP *per capita* which we documented in Table 12 seems to indicate that while the increase in services GDP is largely accounted for by increased productivity, the manufacturing sector is also experiencing an increase in employment. These results are consistent with the anecdotal evidence we discussed in Section 2.2. Municipalities in which oil was discovered became

[34]This is a rough approximation to labor productivity if we assume a Cobb–Douglas production function, for example.

[35]We obtained data on sectoral output per worker for the years 1950, 1960, 1970, 1975, 1980, 1985, 1996, and 2000. Since GDP data is available for 1949 and 1959 but employment data is available for 1950 and 1960, we use the 1949 and 1959 GDP data to get estimates of 1950 and 1960 output per worker, respectively.

local services and commerce hubs for the surrounding area, with these large outfits presenting a significantly higher output per worker than the traditional small-scale service providers.[36]

Summary of Baseline Results. Taken together, our results suggest that local GDP *per capita* and urbanization increase significantly as a result of oil discoveries. While the increase in GDP *per capita* we document is large, the ITT estimates lie within the range estimated for the United States in the literature. Michaels (2011) finds that in the southern U.S. income is 05–28 log points higher in oil-abundant counties than in non-oil counties. He also shows that population density is 30-100 log points higher in oil abundant counties. Allcott and Keniston (2013) look at the impact of resource booms in the U.S. and also find strong results: Resource booms increase both labor income (by about 0.3–0.5 percent points per year during a boom) and employment density (by 60–80 percent) in treated counties. As far as we are aware, there are no previous reliable estimates of the impact of oil discoveries on local economic variables for developing countries. We find that the increase in services GDP is driven by increased output per worker, but the increase in manufacturing GDP must also be driven by an increase in employment.

We do not find a statistically significant increase in population density, but we do document an increase in urbanization.[37] Our sectoral GDP results indicate that oil municipalities might be experiencing a move from rural agricultural activities to provision of service provision in urban areas.[38]

[36]The results for sectoral GDP and output per worker are robust to all of the above robustness exercises but we do not report those tables in the interest of space. Tables are available from the authors upon request.

[37]The result on population density is confirmed if we use overall employment density instead.

[38]Migration as a consequence of oil production in Brazil appears to have been from the countryside to the city within the same MCA rather than from non-oil MCAs to oil MCAs. Historically, inter-municipal migration flows in Brazil tended to be mainly from the northeastern part of the country to the large urban centers in the southeast (S(̃a)o Paulo and Rio de Janeiro), and not within regions (de Lima Amaral (2013)).

5.2 Further results

In this section, we first split discoveries into onshore and offshore and show that, on average, only onshore discoveries seem to have significant positive spillovers. We then use an alternative empirical strategy and estimate a regression which allows us to retrieve the local average treatment effect of oil production. Lastly, we explore the connection between downstream and upstream oil production and show that our results are robust to the exclusion of municipalities with large processing facilities such as refineries or major storage and transportation hubs.[39]

Onshore versus Offshore Discoveries. We distinguish between onshore and offshore discoveries, since some of the channels which we believe can lead to spillovers (such as the physical presence of well-paid oil workers) might be more obviously present for onshore than for offshore locations. In fact, offshore production is concentrated largely of the coast of Rio de Janeiro, and most personnel associated with offshore production is stationed in the municipality of Macaé.

Table 14: Onshore versus Offshore Discoveries 1

	Matched Dry Drilling			
	(1)	(2)	(3)	(4)
VARIABLES	GDP *per cap*	Manufacturing GDP *per cap*	GDP *per cap*	Manufacturing GDP *per cap*
Onshore Discovery Dummy	0.3429***	0.5270**		
	(0.1067)	(0.2157)		
Offshore Discovery Dummy			0.2081	0.4537*
			(0.1315)	(0.2303)
MCA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	768	891	768	891
Number of MCAs	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. The control group is the matched dry-drilling sample.
*** p<0.01, ** p<0.05, * p<0.1

GDP *per capita* in the manufacturing sector increases significantly in both onshore and offshore municipalities. However, when we focus on our measures of spillovers, namely,

[39] In the interest of space, we report results for only our preferred control group (matched dry drilling) from this point on, but as before, all results are very stable across different control groups (and are available upon request).

productivity in the services sector and the urbanization rate, we see that neither of those is affected by offshore discoveries, but there is a large positive impact of onshore discoveries. Labor productivity in the services sector increases by 28%, while the urbanization rate increases by over 5 percentage points (see Tables 14 and 15). The increase in manufacturing GDP shows that offshore discoveries do increase GDP in a mechanical sense. However, we do not find any impact on the local economy. It is also worth noting, however, that the estimated increase in manufacturing GDP is very similar for onshore and offshore discoveries, perhaps indicating that the impact of oil discoveries on non-oil manufacturing is rather limited, even for onshore discoveries.[40]

Table 15: Onshore versus Offshore Discoveries 2

	Matched Dry Drilling			
	(1)	(2)	(3)	(4)
VARIABLES	Services Y/L	Urbanization Rate	Services Y/L	Urbanization Rate
Onshore Discovery Dummy	0.280**	0.0542**		
	(0.135)	(0.0237)		
Offshore Discovery Dummy			0.0187	0.0135
			(0.126)	(0.0313)
MCA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	891	1,024	891	1,024
Number of MCAs	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. The control group is the matched dry-drilling sample. *** p<0.01, ** p<0.05, * p<0.1

Oil Production. We now turn to estimating the impact of oil production rather than oil discoveries on economic outcomes. There are 46 MCAs which have at least one oil production well. As noted earlier, production might be endogenous. We estimate the

[40]While assigning onshore discoveries to municipalities is straightforward, the mapping is not as clear for offshore discoveries (see Section 3.1). To verify whether the offshore result is driven by our measure of offshore discoveries, we used an alternative measure, facing areas, used by the Brazilian Oil and Gas regulator (ANP) to distribute royalties. This is a complex measure, but essentially captures whether a municipality's maritime borders face an oil field (see Monteiro and Ferraz (2012) for a detailed discussion). The resulting measure is substantially broader than ours, since only one MCA can be the one closest to a well, but many MCAs can potentially face it. It thus is ex-ante less likely to pick up spillovers from production. The correlation between the two measures of offshore discoveries is 0.53. We re-ran the regressions using the alternative measure of offshore discoveries, but the results are unchanged.

following equation:

$$Y_{it} = \alpha + \tau_{AT} D_{it} + \beta_t' X_i + \gamma_i + \rho_t + \epsilon_{it}, \tag{2}$$

where we instrument for the production indicator (D_{it}) using our discoveries indicator (Z_{it}) to recover a local average treatment effect. Table 16 qualitatively confirms our ITT results. AS expected, the estimated coefficients are larger. GDP *per capita* increases by over 40%, and urbanization by over 6 percentage points as a consequence of oil production. Similarly, the impact on sectoral GDP is larger.[41] It is intuitive that the ITT results are scaled up by the proportion of compliers. Since the producing municipalities are not a perfect subset of the true discovery municipalities, the instrumental variables specification is not our favourite one and we prefer to report the ITT results as a safe lower bound on the treatment effect.[42]

Table 16: Local Average Treatment Effect of Oil Production

	Matched Dry Drilling					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	Manufacturing GDP *per cap*	Services GDP *per cap*	Agriculture GDP *per cap*
Production Dummy	-0.0190	0.411***	0.0644**	0.725**	0.343**	0.105
	(0.106)	(0.143)	(0.0314)	(0.295)	(0.166)	(0.166)
First Stage F-Stat.	27.38	13.74	27.38	13.33	14.48	13.89
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,024	768	1,024	765	764	765
Number of MCAs	128	128	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. The production dummy is instrumented using the discovery dummy. Discovery is defined as 'True Discovery'.
*** p<0.01, ** p<0.05, * p<0.1

Oil and Gas Processing Production Facilities. For a sample of U.S. counties, Greenstone, Hornbeck, and Moretti (2010) show that there are important local spillovers from the opening of large manufacturing plants. This might also hold true for large downstream oil production facilities such as refineries. Clearly, the decision of where to locate such facilities is likely to be correlated with many unobservables. We therefore do not aim to formally evaluate the impact of downstream production on local economic development, but we want to test whether downstream production facilities are driving most

[41] The same is true for sectoral output per worker (not reported).

[42] We investigate treatment intensity in Appendix C.

of our observed results (as some places with upstream production have also downstream facilities).

Table 17: Excluding Locations with Downstream Production

	Matched Dry Drilling					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	Manufacturing GDP *per cap*	Services GDP *per cap*	Agriculture GDP *per cap*
Discovery Dummy	-0.00430	0.211***	0.0424*	0.455**	0.255**	0.0789
	(0.0730)	(0.0738)	(0.0238)	(0.194)	(0.107)	(0.117)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	904	678	904	676	675	674
Number of MCAs	113	113	113	113	113	113
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as 'True Discovery'.
*** p<0.01, ** p<0.05, * p<0.1

To investigate this possibility we collected data on the location and date of construction of all refineries, directly oil-related factories (such as petrochemical plants) and oil terminals. We also collected data on thermoelectric power plants, which are associated with the oil and gas industry.[43] We observe that discoveries increase the probability of hosting a downstream facility by roughly 10%, which is not negligible but not overwhelming either. This rises to 15% when we use an ad-hoc measure for large discoveries (top 20 in the year 2000 in terms of number of discovery wells).

There is thus some support for the hypothesis that discoveries tend to lead to the establishment of downstream production facilities in an MCA. To evaluate the pure impact of the upstream sector, we thus exclude those municipalities which host a downstream production facility from both the treatment and the control group and then re-estimate our baseline specification. As can be seen by comparing Table 17 with Tables 8 and 12, the results do not appear to be driven by downstream production facilities only. Upstream oil production thus directly impacts the local economy, even when it generates no significant royalties and does not lead to the establishment of downstream production facilities.

[43]See Appendix A for details on data construction.

5.3 Explaining the Mechanism

In this section we investigate the mechanisms underlying our results in more detail. We aim to shed light on three questions: (i) Are the services sector results driven mainly by local price effects or real economic change? (ii) What happens to non-oil manufacturing? and (iii) What happens to the agricultural sector?

Due to constraints on the availability of microdata, this more in-depth analysis cannot be conducted using our preferred difference-in-difference identification strategy. Thus we exploit a cross-sectional identification. We use matched worker–firm microdata: the Ministry of Labor's RAIS (Relação Anual de Informações Sociais). The RAIS dataset has information on each formal worker at each plant in Brazil. Since RAIS looks only at formal workers we complement this data with data on informality from the 2000 population census, collected by the Brazilian Bureau of Statistics. We use cross-sectional data for the year 2000, because this is the first year for which high-quality data from both the employment and population censuses are available.[44]

To guarantee maximum comparability with the results reported in previous sections of the paper, we use the same assigned to treatment and control groups. In terms of the identification, we showed in Tables 3, 5 and 6 (see Section 3.3) that drilling attempts depend on geology and are not correlated with MCA characteristics at the time of drilling. Given that discoveries are random (conditional on drilling) even a cross-sectional comparison of treatment and control groups allows for some insights into at least the qualitative impact of oil discoveries. We estimate the following equation:

$$Y_i = \alpha + \tau_{cs} Z_i + \beta' X_i + \epsilon_i, \tag{3}$$

where Y_i is the outcome variable in 2000, X_i includes the usual controls, and Z_i equals 1 if oil was discovered in the MCA unit between 1940 and 2000.

Table 18 shows the baseline results from the cross-sectional exercise. The first three columns confirm the previous findings: In 2000, the assigned to treatment group has a

[44] See Appendix D for more information on the microdata.

Table 18: Oil Discoveries, Wages, Worker Density, and Firm Density

	Matched Dry Drilling					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Population Density	ln GDP *per capita*	Urbanization Rate	ln Worker Density	ln Firm Density	ln Average Wage
Discovery Dummy	-0.0269	0.396***	0.0551*	0.506*	0.384	0.185**
	(0.129)	(0.120)	(0.0301)	(0.287)	(0.285)	(0.0739)
Observations	128	128	128	128	128	128
Number of MCAs	128	128	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS

Notes. Robust standard errors parentheses. Discovery is defined as "True Discovery". Cross-sectional estimation using microdata for the year 2000. Densities are specified as the number of firms and workers per square kilometer.
*** p<0.01, ** p<0.05, * p<0.1

higher *per capita* GDP and is more urbanized, but population density is not affected by oil discoveries.[45] Additionally, Columns (4)–(6) show that places where oil was discovered have higher average wages and a higher worker density, but firm density is not statistically different between discovery and control groups.[46] MCAs where oil was discovered are thus richer and more urbanized, pay higher wages, and have more formal workers. To investigate which sectors are affected by oil discoveries, we construct sectoral measures of firm and worker density. Importantly, we are able to exploit subsector identifiers in the microdata to obtain a manufacturing sector without extractive activities, which was not possible using the historical data.

Table 19: Oil Discoveries, Worker Density and Firm Density by Sector

	Matched Dry Drilling					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Manufacturing Firm Density	ln Services Firm Density	ln Agriculture Firm Density	ln Manufacturing Worker Density	ln Services Worker Density	ln Agriculture Worker Density
Discovery Dummy	0.308	0.426	0.274	0.338	0.796**	0.546
	(0.338)	(0.302)	(0.286)	(0.450)	(0.353)	(0.359)
Observations	128	128	128	128	128	128
Number of MCAs	128	128	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS

Notes. Robust standard errors in parentheses. Discovery is defined as "True Discovery". Cross-sectional estimation using microdata for the year 2000. Densities are specified as the number of firms and workers per square kilometer.
*** p<0.01, ** p<0.05, * p<0.1

Table 19 highlights that the manufacturing sector (excluding natural resource extrac-

[45]We focus on the matched subsample in this section but results are unchanged when using the full sample of dry drilling MCAs as the control group.

[46]Densities are specified as the number of firms and workers, respectively, per square kilometer.

tion) and the agricultural sector are not affected by oil production. We do not find any evidence for a Dutch-disease style crowding-out of the manufacturing sector nor of positive spillovers from oil production to manufacturing. By contrast, the growth in the number of formal workers is driven by an increase in the number of formal workers in services.

Table 20 further disaggregates the data for services (where we verified an impact of oil discoveries). First, we observe that average firm size in the services sector is significantly higher in the assigned to treatment group. We know from the labor literature (see Idson and Oi (1999), for example) that larger establishments tend to be more productive, and this could be a driver for development. Secondly, the numbers of both skilled and unskilled workers in services is higher in oil MCAs, but while the average skilled wage is also significantly higher, the unskilled wage is not affected.[47] An interesting picture thus emerges. In municipalities in which oil was discovered, more workers are employed in the services sector, services firms are larger, and the skilled workers in the services sector receive higher wages. In other words, the local services sector grows with oil discoveries. The fact that the skilled wage is higher but the unskilled wage is not points to differences in the supply curve for skilled and unskilled workers. The elasticity for unskilled workers appears to be so high that more workers can be attracted at virtually no higher pay, while the supply of skilled workers is relatively more inelastic in comparison.

An interesting question is where the new services workers are drawn from. Neither population density increases, nor does formal employment density in non-services sectors decrease. In other words, there is no significant in-migration, and no sectoral relocation of formal workers. While we cannot rule out that there are changes which are on average too small for us to detect, it seems unlikely that these can fully explain the "new services workers". What appears more likely is that they are mainly drawn from the informal sector.

In our sample, the informal sector is very large: On average, only 35% of workers are formally employed and only 25% have a valid work card.[48] Table 21 uses population

[47] Skilled workers are defined as those who at least completed high school.

[48] The definition of formal employment is from the Brazilian Bureau of Statistics and includes workers

Table 20: Oil Discoveries and the Services Sector

	Matched Dry Drilling					
	All Dependent Variables for Services Sector					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ln Firm Size	ln Skilled Worker Density	ln Unskilled Worker Density	Skilled Worker Fraction	ln Avg. Skilled Wage	ln Avg. Unskilled Wage
Discovery Dummy	0.370***	0.711*	0.685**	-0.0188	0.168**	0.0860
	(0.118)	(0.363)	(0.346)	(0.0260)	(0.0793)	(0.0611)
Observations	128	128	128	128	128	128
Number of MCAs	128	128	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS

Notes. Robust standard errors in parentheses. Discovery is defined as "True Discovery". Cross-sectional estimation using microdata for the year 2000. Densities are specified as the number of firms and workers per square kilometer. Skilled workers are defined as those who at least completed high school.
*** p<0.01, ** p<0.05, * p<0.1

census data to show that oil discoveries are associated with a larger fraction of workers employed in the formal sector. The higher formalization rate offers an explanation for where the additional workers in the services sector come from: They move from the informal to the formal sector. Since the pool of workers in the informal sector tends to be predominantly unskilled, this also explains the higher elasticity of labor supply for unskilled workers.

We also check whether labor force participation increases and the fraction of self-employed workers decreases with oil discoveries (as workers from low productivity self-employed services provision move to larger formal services firms, for example). As shown in Columns (3) and (4) of Table 21 we find evidence for a decline in self-employment but no evidence for a higher labor force participation rate.

To gauge from which informal sector workers move to the formal services sector, we use the population census and decompose the overall workforce into broad categories. Column (5) of Table 21 confirms that in the discovery group a significantly larger fraction of the overall workforce is employed in extractive industries than in the control group. Recall that we showed that the density of *formal* employees in the agricultural sector does not differ between treatment and control group. However, column (6) of Table 21 shows that the overall fraction of workers in the agricultural sector is significantly lower in

with a valid work card, those who work in the military or judiciary, and self-employed workers who contribute to social security.

Table 21: Oil Discoveries and Labor Informality

	Matched Dry Drilling							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Percentage of Workers in the Formal Sector	Percentage of of Workers with Valid Employment Card	Labor Force Participation Rate	Percentage of Workers who are Self-Employed	Percentage of Workers Employed in Extractive Industries	Percentage of Workers Employed in Agriculture	Percentage of Workers Employed in Manufacturing	Percentage of Workers Employed in Services
Discovery Dummy	4.352**	4.481***	0.0660	-2.627**	0.579***	-5.364**	-0.0206	4.452**
	(1.710)	(1.635)	(0.850)	(1.316)	(0.214)	(2.632)	(0.742)	(2.191)
Observations	128	128	128	128	128	128	128	128
Number of MCAs	128	128	128	128	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Notes. Robust standard errors in parentheses. Discovery is defined as "True Discovery". Cross-sectional estimation using microdata for the year 2000. Formal employment includes workers with a valid work card, those who work in the military or judiciary, and self-employed workers who contribute to social security.
*** p<0.01, ** p<0.05, * p<0.1

municipalities which discovered oil. The number of *informal* workers in agriculture must therefore be lower. Columns (7) and (8) confirm the earlier results for the manufacturing and services sectors, that is, no impact on employment in the manufacturing sector and an increase in employment in the services sector.

Brazil still had a large subsistence farming sector in 2000, which employed a substantial number of people with very low productivity. In places with oil discoveries, we can observe a move of these informal agricultural workers to an expanding services sector. Our results indicate that this is the main positive externality from oil discoveries. Overall, places which discovered oil have larger, more productive services sectors, probably driven by an increase in local demand for non-tradables from oil workers and the oil-producing firms. The increased demand for labor leads to more workers being pulled into the services sector and an increase in the wage of skilled workers. The unskilled wage does not increase, as there is ample supply of unskilled workers in the informal agricultural sector. This move from rural informal work to the formal sector in urban areas also explains the observed increase in urbanization.

It is worth noting that no impact on the manufacturing sector was found. This might be somewhat specific to the particular situation of a developing country with relatively little large-scale manufacturing in the affected regions. The impact of oil discoveries on wages of skilled workers in the services sector hinted at the possibility that in locations lacking an ample supply of labor in the informal, subsistence agricultural sector, an upward-sloping labor supply curve would drive up manufacturing wages and potentially lead to the local Dutch-disease type effects often hypothesized. On the other hand, positive technological spillovers from oil production might also exist in regions where there is an important nucleus of high-end manufacturing. In the Brazilian case, the presence of an oil sector and the associated increase in local demand for non-tradables had a strong impact on the development of the local services sector and precipitated a decrease in the highly unproductive subsistence farming sector and thus furthered local economic development.

6 Conclusion

We investigated the effects of natural resource extraction on local economic development and documented a positive growth effect of oil discoveries. We found a positive impact of oil discoveries on urbanization as well as increases in services GDP, services output per worker, and the size of services firms. We did not find evidence of de-industrialization in oil municipalities. By comparing municipalities where drilling turned up dry wells to those where oil was discovered, we constructed a unique control group based on random assignment. Since we examined the entire track of oil discoveries in Brazil, we were able to provide evidence that there were no pre-treatment differences between our treatment and control groups.

It is important to highlight that our results apply to a specific institutional framework, given that we studied the effects of oil discoveries on the local development of only one country. For instance, the U.S. has a more widespread ownership of resources than Brazil. There are thousands of oil companies in the U.S., in contrast to the historical monopoly of Petrobras in Brazil. Because of this market structure, oil services are more likely to be concentrated in just a few places in Brazil. By contrast, in the U.S. an entire chain of small oil services can be located close to the more widespread oil firms. Finally, we cannot rule out the possibility that oil discoveries positively affect local development of oil municipalities but have adverse effects at the national level (through, for example, a nominal appreciation). We show that at the local level, oil discoveries are not a curse per se, and the pure market effect (i.e., in the absence of any fiscal windfalls) benefits development. In light of the results on fiscal windfalls in the literature, it appears that the impact of the windfall effect of resource wealth is strongly dependent on the institutional setting. While natural resource extraction can foster local growth, defining good policies and institutions for use of the associated fiscal windfalls thus remains a key policy challenge.

References

Acemoglu, D., and M. Dell (2010): "Productivity Differences between and within Countries," *American Economic Journal: Macroeconomics*, 2(1), 169–88.

Acemoglu, D., A. Finkelstein, and M. J. Notowidigdo (2014): "Income and Health Spending: Evidence from Oil Price Shocks," *Review of Economics and Statistics*, Forthcoming.

Allcott, H., and D. Keniston (2013): "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America," mimeograph, Yale University.

Angrist, J. D., and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), pp. 431–442.

ANP (2001): "Guia dos Royalties do Petróleo e do Gás Natural," Discussion paper, Agencia Nacional do Petróleo, Gás natural e Biocombustíveis.

——— (2012): "Anuário Estatístico Brasileiro do Petróleo, Gás natural e Biocombustíveis," Anuário estatístico da anp, Agencia Nacional do Petróleo, Gás natural e Biocombustíveis.

Arezki, R., and M. Brueckner (2011): "Oil Rents, Corruption, and State Stability: Evidence from Panel Data Regressions," *European Economic Review*, 55(7), 955 – 963.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should We Trust Differences-In-Differences Estimates?," *The Quarterly Journal of Economics*, 119(1), 249–275.

Bosco, F. (2003): "50 anos da Petrobras," *Petro&Química*, 252(1).

Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini (2013): "The Political Resource Curse," *American Economic Review*, 103(5), 1759–96.

Caselli, F., and G. Michaels (2013): "Do Oil Windfalls Improve Living Standards? Evidence from Brazil," *American Economic Journal: Applied Economics*, 5(1), 208–38.

Caselli, F., and A. Tesei (2011): "Resource Windfalls, Political Regimes, and Political Stability," Working Paper 17601, National Bureau of Economic Research.

CNI (2012): "A contribuição do setor brasileiro de petróleo, gás e biocombustíveis para o desenvolvimento sustentável no país," Cadernos setoriais rio+20, Confederação Nacional da Indústria.

Cust, J., and T. Harding (2014): "Institutions and the Location of Oil Exploration," Discussion Paper 127, OxCarre Research Paper.

Da Mata, D., U. Deichmann, J. Henderson, S. Lall, and H. Wang (2007): "Determinants of city growth in Brazil," *Journal of Urban Economics*, 62(2), 252–272.

Davis, D. R., and D. E. Weinstein (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92(5), 1269–1289.

de Lima Amaral, E. F. (2013): *Brazil: internal migration.*Blackwell Publishing Ltd.

Dube, O., and J. Vargas (2013): "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *The Review of Economic Studies*, Forthcoming.

Fahim, M., T. Al-Sahhaf, and A. Elkilani (2009): *Fundamentals of Petroleum Refining*. Elsevier Science.

Frölich, M., and M. Lechner (2010): "Exploiting Regional Treatment Intensity for the Evaluation of Labor Market Policies," *Journal of the American Statistical Association*, 105(491), 1014–1029.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118(3), 536–598.

Harbaugh, J., J. Davis, and J. Wendebourg (1995): *Computing Risk for Oil Prospects: Principles and Programs*, Computer Methods in the Geosciences. Elsevier Science.

Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007): "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15(3), 199–236.

Hyne, N. (2001): *Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production*, PennWell nontechnical series. PennWell Corporation.

Idson, T. L., and W. Y. Oi (1999): "Workers Are More Productive in Large Firms," *American Economic Review*, 89(2), 104–108.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Isham, J., M. Woolcock, L. Pritchett, and G. Busby (2005): "The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth," *The World Bank Economic Review*, 19(2), 141–174.

Kellogg, R. (2011): "Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch," *The Quarterly Journal of Economics*, 126(4), 1961–2004.

Krugman, P. (1987): "The narrow moving band, the Dutch disease, and the competitive consequences of Mrs. Thatcher : Notes on trade in the presence of dynamic scale economies," *Journal of Development Economics*, 27(1-2), 41–55.

Kuralbayeva, K., and R. Stefanski (2013): "Windfalls, structural transformation and specialization," *Journal of International Economics*, 90(2), 273 – 301.

Michaels, G. (2011): "The Long Term Consequences of Resource-Based Specialisation," *Economic Journal*, 121(551), 31–57.

Mohn, K., and P. Osmundsen (2008): "Exploration economics in a regulated petroleum province: The case of the Norwegian Continental Shelf," *Energy Economics*, 30(2), 303 – 320.

Monteiro, J., and C. Ferraz (2012): "Does Oil Make Leaders Unaccountable? Evidence from Brazil's offshore oil boom," mimeograph, PUC-Rio.

Moretti, E. (2011): *Handbook of Labor Economics*vol. 4b, chap. Local Labor Markets, pp. 1237–1313. New York: Elsevier.

Petroleum Extension Service (1997): *Fundamentals of Petroleum*. Petroleum Extension Service, University of Texas at Austin.

——— (2005): *A Dictionary For The Oil And Gas Industry*. Petroleum Extension Service, University of Texas at Austin.

Reis, E., P. Tafner, M. Pimentel, R. Serra, L. Reiff, K. Magalhaes, and M. Medina (2004): "Estimativas do PIB dos Municípios Brasileiros, 1970-1996: Metodologia e Resultados," mimeograph, Ipea.

Rosenbaum, P. R., and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

Rubin, D. (2001): "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services and Outcomes Research Methodology*, 2(3-4), 169–188.

SACHS, J., AND A. WARNER (1995): "Natural Resource Abundance and Economic Growth," *NBER Working Paper 5398.*

STEFANSKI, R. (2010): "Structural Transformation and the Oil Price," OxCarre Working Papers 048, Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.

VAN DER PLOEG, F. (2011): "Natural Resources: Curse or Blessing?," *Journal of Economic Literature*, 49(2), 366–420.

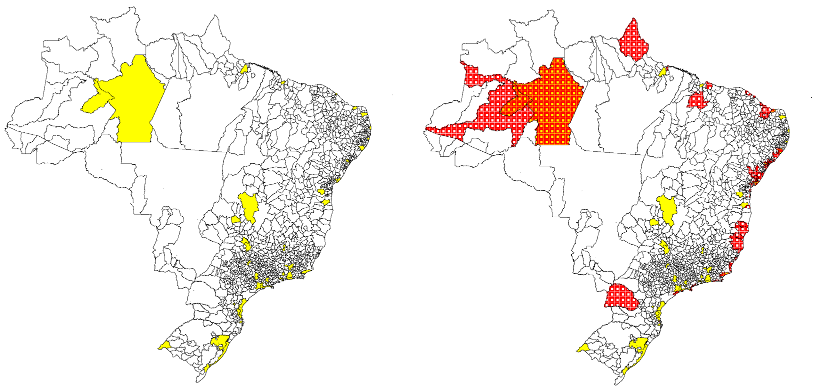
YERGIN, D. (2008): *The Prize: The Epic Quest for Oil, Money & Power*. Simon & Schuster.

NOT FOR PUBLICATION

A Appendix

A.1 Downstream Production Data

Fig. A.1: Processing Facilities



(a) Processing Facilities (Yellow) (b) Discovery (Red), Facilities (Yellow)

Notes. The figures show 1,275 Minimum Comparable Areas (MCAs) in 1940. The discovery dummy is the "All Discoveries" dummy (which is equal to one when at least one field, subfield, or field extension discovery was made in the MCA).

Here we discuss the role of the downstream industry in Brazil (processing and transportation facilities). Information on the construction date of each refinery, and of each onshore or offshore terminal, is from Petrobras and Transpetro. Information on the construction dates of petrochemical plants and thermoelectric power plants is from Petrobras and various online sources. By the year 2000 there were 15 refineries or directly oil-related factories, 18 onshore oil terminals, 22 offshore terminals, and 2 thermoelectric power plants in Brazil. Using these data we constructed an indicator which we set equal to 1 if an MCA has at least one of those oil-related processing or transportation facilities. To evaluate the link between the upstream and downstream oil sectors, we regress the production facilities dummy on the indicator for "True Discoveries". As before, a full set of controls is included. We again include MCA and year fixed effect and clustered standard errors at the MCA level. Regardless of the control group, the coefficient for the true discoveries dummy is positive and significant.

We also collected data provided by ANP (2001) detailing which municipalities they classify as the main production and main production support sites, respectively. The idea is to perform an additional test of the hypothesis that production facilities are more likely to be located in MCAs in which large reserves of oil were discovered. Main production sites are defined as locations with facilities for processing, treating, storing, and transporting oil. Support sites are those with ports, airports, heliports, offices, or similar facilities used to support the extraction, production, and processing of oil. We match these municipal data to the relevant MCAs and then construct a new indicator at the MCA level. Unfortunately, the data are available for only the year 2000, so we do not

Table A.1: Discoveries and Processing Facilities

	Matched Dry Drilling	
	(1)	(2)
VARIABLES	Production Facilities Dummy	Production Facilities Dummy
Discovery Dummy	0.102**	
	(0.0486)	
Large Discovery Dummy		0.147**
		(0.0709)
MCA FE	Yes	Yes
Year FE	Yes	Yes
Observations	896	896
Number of MCAs	128	128
Geographical Controls	Yes	Yes
Initial Conditions	Yes	Yes
Estimation	FE	FE

Notes. Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. The initial conditions with time-varying coefficients are GDP *per capita* in 1949, urbanization rate in 1940, and population density in 1940. The geographical controls with time-varying coefficients are latitude and longitude, the Amazon dummy, and the coastal dummy. Discovery is defined as "True Discovery". "Large Discovery" is a discovery which makes the hosting municipality one of the top 20 in terms of number of wells.
*** p<0.01, ** p<0.05, * p<0.1

know the year in which the various municipalities first became main production or support sites. Therefore, we are unable to use these variables in a panel regression. Nevertheless, it is worth pointing out that the correlations between having had a discovery and being a main production or main production support site are 0.2466 and 0.2747, respectively.

A.2 Comparison of Treatment and Control Groups

The normalized difference (ND) for continuous variables is given by

$$ND = \frac{\mu_t - \mu_c}{\sqrt{\sigma_t^2 + \sigma_c^2}},$$

where μ_t and σ_t^2 are the mean and variance of the treated group, and μ_c and σ_c^2 are the corresponding values for the control group.

The ND for dichotomous variables is defined as

$$ND = \frac{p_t - p_c}{\sqrt{p_t(1-p_t) + p_c(1-p_c)}},$$

where p_t and p_c are the proportions (prevalences) for the treated and control groups, respectively.

Imbens and Wooldridge (2009) suggest that for a standardized difference of more than 0.25, "linear regression methods tend to be sensitive to the specification" (p. 24). Table A.2 systematically investigates the differences between the assigned to treatment and control groups using standardized differences. As can be seen there, the matched dry-drilling MCAs constitute a good control group in terms of observables.

Table A.2: Overlap of Treated and Various Control Groups

Variable		(I) Oil Discovery	(II) Dry Drilling	(III) Matched Dry Drilling	(IV) No Discovery in Oil States	(V) Neighbors	(VI) Matched Neighbors	(VII) Dry No Neighbor
Pop. Density 1940	Mean	32.89	30.33	35.09	30.15	24.54	31.74	35.2
	S.D.	51.35	132.29	104.47	78.22	50.13	72	153.1
	Standardized Difference	-	0.018	-0.019	0.029	0.116	0.013	-0.014
Urbanization 1940	Mean	0.27	0.22	0.24	0.21	0.18	0.21	0.23
	S.D.	0.18	0.18	0.2	0.15	0.14	0.16	0.19
	Standardized Difference	-	0.196	0.111	0.256	0.395	0.249	0.153
GDP per capita 1949	Mean	0.67	0.88	0.69	1.13	0.61	0.57	1.06
	S.D.	0.42	0.89	0.75	0.98	0.5	0.5	1
	Standardized Difference	-	-0.213	-0.023	-0.431	0.092	0.153	-0.360
Manufacturing/GDP 1949	Mean	0.19	0.13	0.13	0.13	0.13	0.13	0.11
	S.D.	0.15	0.16	0.17	0.14	0.15	0.15	0.12
	Standardized Difference	-	0.274	0.292	0.292	0.283	0.283	0.416
Services/GDP 1949	Mean	0.38	0.37	0.4	0.36	0.34	0.36	0.37
	S.D.	0.2	0.21	0.23	0.18	0.2	0.22	0.2
	Standardized Difference	-	0.034	-0.066	0.074	0.141	0.067	0.035
Agriculture/GDP 1949	Mean	0.43	0.51	0.48	0.52	0.55	0.53	0.52
	S.D.	0.24	0.24	0.26	0.23	0.25	0.27	0.23
	Standardized Difference	-	-0.236	-0.141	-0.271	-0.346	-0.277	-0.271
Altitude	Mean	78.81	229.15	143.38	384.27	179.48	109.39	276.4
	S.D.	97.96	247.65	212.68	273.19	206	104.12	259.1
	Standardized Difference	-	-0.565	-0.276	-1.053	-0.441	-0.214	-0.713
Avg. Rainfall	Mean	118.46	127	122.23	108.34	121.78	118.78	120.9
	S.D.	38.79	43.65	51.44	36.96	49.24	47.1	37.63
	Standardized Difference	-	-0.146	-0.059	0.189	-0.053	-0.005	-0.045
Avg. Temperature	Mean	24.95	23.96	24.35	22.9	24.28	24.8	23.42
	S.D.	1.9	2.97	2.7	2.91	2.72	2.16	3.06
	Standardized Difference	-	0.281	0.182	0.590	0.202	0.052	0.425
Latitude	Mean	-11.88	-13.72	-12.62	-15.85	-12.03	-11.49	-15.8
	S.D.	6.44	9.67	8.6	8.05	8.27	7.47	9.72
	Standardized Difference	-	0.158	0.069	0.385	0.014	-0.040	0.336

Continued on next page

Table A.2 – *Continued from previous page*

Variable		**(I)** Oil Discovery	**(II)** Dry Drilling	**(III)** Matched Dry Drilling	**(IV)** No Discovery in Oil States	**(V)** Neighbors	**(VI)** Matched Neighbors	**(VII)** Dry No Neighbor
Longitude	Mean	-40.65	-46.94	-43.5	-44.53	-44.32	-42.65	-46.83
	S.D.	6.46	7.31	7.6	5.18	8.46	8.33	5.42
	Standardized Difference	-	0.645	0.286	0.469	0.345	0.190	0.733
Coastal Indicator	Prop.	0.59	0.3	0.53	0.11	0.19	0.42	0.29
	Standardized Difference	-	0.431	0.086	0.823	0.636	0.244	0.448
Semiarid Indicator	Prop.	0.19	0.15	0.23	0.25	0.25	0.28	0.13
	Standardized Difference	-	0.075	-0.070	-0.103	-0.103	-0.151	0.116
Amazon Indicator	Prop.	0.08	0.3	0.17	0.1	0.23	0.15	0.25
	Standardized Difference	-	-0.413	-0.194	-0.049	-0.300	-0.156	-0.333
	Number of MCAs	64	158	64	711	156	64	104

Note: Oil discovery is the treated group of 64 MCAs. Six control groups are shown: MCAs where drilling took place but nothing was found (column II: "Dry Drilling"), propensity score matched sample of MCAs where drilling took place but nothing was found (column III: "Matched Dry Drilling"), MCAs with no oil discovery but in states where oil was discovered in other MCAs (column III: "No Discovery in Oil States", MCAs that are adjacent to the treated MCAs (column IV: "Neighbors"), propensity score matched sample of MCAs that are adjacent to the treated MCAs (column VI: "Matched Neighbors"), and MCAs where drilling took place but nothing was found and which are not adjacent to treated MCAs (column VII: "Dry, No Neighbour"). Dry drilling and matched dry drilling are our baseline control groups. We use the other four to study robustness. The last row shows the total number of MCAs in each group. Green entries indicate standardized difference with values up to 0.2. Orange entries indicate standardized difference with values between 0.2 and 0.3. Red entries indicate standardized difference with values greater than 0.3.

A.3 Summary Statistics

Table A.3: Summary Statistics: Minimum Comparable Areas

Category	Variable	Mean	Std. Dev.	Min.	Max.	N
Outcome Variables	Urban Population/Total Population	0.458	0.253	0.015	1	10,197
	Log of Population Density	3.199	1.316	-3.222	9.186	10,198
	Log of GDP per capita	0.501	0.985	-4.602	6.38	7,645
	Share of GDP: Manufacturing	0.195	0.169	0	0.971	11,436
	Share of GDP: Services	0.431	0.171	0.001	0.975	11,443
	Share of GDP: Agriculture	0.362	0.232	0	1	11437
Oil Variables	All Discoveries Dummy	0.024	0.151	0	1	77,775
	Oil Production Dummy	0.017	0.131	0	1	77,775
	True Discoveries Dummy	0.016	0.125	0	1	77,775
	Stock of Producer Wells	2.47	35.322	0	1814	77,775
	Stock of Discovery Wells	0.371	4.761	0	218	77,775
	Stock of Injection Wells	0.252	4.078	0	131	77775
Geography	Average Altitude	439.119	303.067	0	1278	77,775
	Average Temperature	22.669	2.841	14.965	27.88	77,775
	Average Rainfall	109.93	34.287	34.63	258.358	77,775
	Indicator: Amazon Region	0.073	0.26	0	1	77,775
	Indicator: Semiarid Region	0.231	0.422	0	1	77,775
	Indicator: Coastal MCA	0.107	0.309	0	1	77,775
Pre-Treatment Variables	Log of Population Density in 1940	2.701	1.305	-3.228	7.562	77,714
	Urbanization Rate in 1940	0.219	0.154	0	1	77,775
	Log of GDP per capita in 1949	-0.326	0.854	-4.602	1.828	77,653

Notes. Data from ANP (Brazilian oil and gas industry regulator) and Ipeadata. Data aggregated and treated for 1,275 Minimum Comparable Areas (MCAs). The total number of observations is the product of the number of MCAs and the number of years in our sample (during the period 1940–2000). Temperature is measured in degrees Celsius, precipitation in millimeters per month, and altitude in meters.

B Appendix: More Details on Oil Extraction and the Institutional Background

B.1 On Oil Drilling and Production

There is an extensive literature on the principles and practices of oil drilling and production (e.g., from petroleum geology and petroleum engineering). In this appendix, we aim to clarify selected aspects of drilling and production that are relevant to our research design, without detailing every single aspect of oil (and gas) exploration and production.

Oil exploration and production are associated with risk. Although there are several technical methods for appraising hydrocarbon resources, the industry always works with limited information on the existence of hydrocarbon deposits. The uncertainty is related to the location, volume, and quality of hydrocarbon deposits. Even with enough geological information, there is always the risk of drilling a dry exploratory hole or not discovering commercial quantities of oil. There are also risks during the production phase, such as the price of oil, costs and taxes, institutional uncertainty, regulation, natural disasters, and accidents. Offshore drilling in deep water presents even greater challenges. According to Harbaugh, Davis, and Wendebourg (1995), luck is a major factor in oil exploration. The name for an exploratory well (called a "wildcat") speaks for itself regarding the inherent risk of the oil business.

The petroleum industry is loosely divided into two segments: upstream and downstream. The upstream industry comprises exploration and production activities, where production activities means the process of recovering petroleum from the subsurface. Upstream activities occur both onshore and offshore. The downstream industry entails processing, retailing, and transporting of petroleum.

Oil exploration involves several steps and uses a compilation of knowledge from geology, geophysics, and geochemistry. The oil company aims to find an oil field — a contiguous geographic area with oil. First, petroleum professionals collect useful geological information on a "prospect" (a delimited area that possesses certain geological features that may induce drilling). By "useful information," they mean a source rock, a reservoir, and a trap.[49] A *source rock* is a rock within which oil or gas is generated from organic material (Petroleum Extension Service (2005)). A source rock is usually a shale rock. Nevertheless, not every shale has enough biogenic material to be classified as a source rock. A *reservoir* accumulates hydrocarbons and is made from porous rocks. Porosity is needed for the accumulation of hydrocarbons, and basically only sedimentary rocks are porous enough. Typical sedimentary rocks that form a reservoir include sandstone and limestone. The "quality" of the oil inside the reservoir can vary, depending

[49] There are three types of rock according to how they are formed: igneous (from magma), sedimentary (from erosion), and metamorphic (a heated sedimentary or igneous rock). Sedimentary rocks are of great interest in regard to oil production, because they are the ones in which petroleum accumulation chiefly occurs. An example of a sedimentary rock is shale, which originates as clay and is compacted by subsurface pressure and weight. Other examples of sedimentary rock are sandstone (from sand) and limestone (from shells).

on its properties and impurities (e.g., the presence of sulfur and metals). The company also looks for areas with specific geological features called *traps*. A hydrocarbon trap is composed of two elements: a structure (subsurface contortion) and a seal. Hydrocarbon molecules are lighter than water, and there are subsurface contortions that induce the hydrocarbons to move upward towards the surface (e.g., anticlines and faults). Therefore, there is a need for a "seal" to prevent the hydrocarbons from spilling out onto the surface. A seal is a rock with low permeability. (Just as porosity is important for the accumulation of hydrocarbons in a reservoir, the degree of connections between pores in a rock formation determines whether a rock can serve as a seal.) Although shale has high porosity, it also has very low permeability, and typically makes for a good seal.

In sum, the area should possess certain characteristics, such as abundant sandstone for reservoir rocks, shale for hydrocarbon source rocks, and numerous geological structures for potential trapping of hydrocarbons. Each oil field has its own "fingerprint" and its unique characteristics lead to a case-by-case analysis of drilling attempts. Wells are very expensive to drill, so studies in advance of a decision to drill must be as accurate and precise as possible.

If there are strong indications of potentially oil-bearing formations, the oil company may drill an exploratory well. Even with all positive indications of oil, only by *making a hole* can the company be sure of the presence (or absence) of oil. During the drilling process, data acquisition is key. There are several "logging-while-drilling" (information-recording) procedures that are performed during the drilling phase, for example, to differentiate between permeable and impermeable rock formations.[50] Depending on the outcome of the exploratory drilling, the company evaluates the well's hydrocarbon potential. Not even evidence of hydrocarbon deposits as indicated in the logs is a guarantee that oil production is feasible. One can assign *a priori* probabilities before drilling, and then revise the probability of success given the result of the drilling attempt. Updated probabilities can be used as a source of knowledge to be applied to future drilling attempts. Depending on the preliminary information accumulated during drilling, the well can be abandoned or not. In the end, using all information available the company decides whether the drilling generated a discovery or a *dry hole*.

After a discovery, the appraisal continues: Additional drilling is required to delineate the size and extension of the oil field.[51] "Step-out" wells (delineation or appraising wells) are the wells used to evaluate the extent of the field. The more that is known about the oil field, the easier and less expensive to drill additional wells. Generally, the number of step-out wells is positively correlated with the magnitude of the field that was discovered. Once the oil company has delineated the oil field and is confident about the viability of production, it undertakes to *complete* the existing wells and drill additional production wells (producer wells). To complete a well means to perform the necessary operations to bring fluids to the surface (Petroleum Extension Service (1997)). After completion and the drilling of producer wells, the oil and gas production cycle begins. Production cycle occurs after exploration has proven successful. An economic assessment of the production cycle should entail reserve and risk calculations (Hyne (2001)).

The production cycle involves a natural phase and an enhanced phase. Initially, natural pressure in the reservoir brings oil from the reservoir to the surface. Normally, the

[50] One example is the logging of information on the cuttings of rock brought to the surface by the circulating drilling fluid.

[51] "Play" is the name used to describe the extent of a hydrocarbon-bearing formation.

reservoir pressure falls as production proceeds, and the oil company needs to (artificially) increase the pressure in the well. The addition of artificial pressure to optimize production is broadly called "enhanced oil recovery" and is divided into primary, secondary, and tertiary recovery. *Primary recovery* (or primary production) means to use an artificial method of lifting. The most common artificial lift system is a *beam pump* to pump up the oil. During primary recovery, only a small percentage of the hydrocarbon deposits are extracted. *Secondary recovery* aims to restore the reservoir pressure by injecting water (waterflooding) or gas. Secondary recovery is costly, because it requires huge amounts of water or gas. To boost the pressure once it begins to fall, new wells are drilled (injection wells) for injection of water and gas, usually at the edges of the oil field. This injection aims either to slow a decline in production or to increase production. Finally, *tertiary recovery* is done by injection of steam or special chemicals (chemical flooding) into the reservoir. In practice, all three recovery phases can occur concurrently.[52]

Enhanced recovery is so important in the petroleum industry that the location of the producer well is chosen with the secondary well (injection well) in mind. As mentioned before, efforts to enhance recovery are costly and are dependent upon the state of the economy and the potential oil recovery volume. Consequently, repeated monitoring of a reservoir is essential to choose the best locations for the injection wells. The idea is to design an optimal distribution of injection wells so as to optimize long-term production.

There are several types of wells: wildcat well, rank wildcat well, step-out well, producer well, injection well, etc. Since there are different steps in the process of obtaining oil, wells are classified broadly as exploratory wells and development wells. Examples of *exploratory wells* are wildcat wells (drilled a mile or more from an area of existing oil production) and rank wildcat wells (drilled in an area where there is no existing production). If the exploratory drilling proves successful, the company starts to drill step-out wells (also included in the exploratory well category). After the oil field has been delineated, the company starts to drill production wells within the known extent of the field. Every well drilled inside the known extent of the field is called a *development well* (Hyne (2001)). The development well category includes producer wells and injection wells (recall that injection wells are drilled to enhance oil recovery). Different categories of wells have different probabilities of finding oil. On average, rank wildcat exploratory wells have lower success ratio than step-out wells. An oil company can rank wells in terms of probability, even in the face of uncertainty. The American Petroleum Institute reported that in 2000 the success rate for wildcat wells was 39% (Hyne (2001)). Note that an unsuccessful drilling is classified as a dry hole in both exploratory and development well categories.

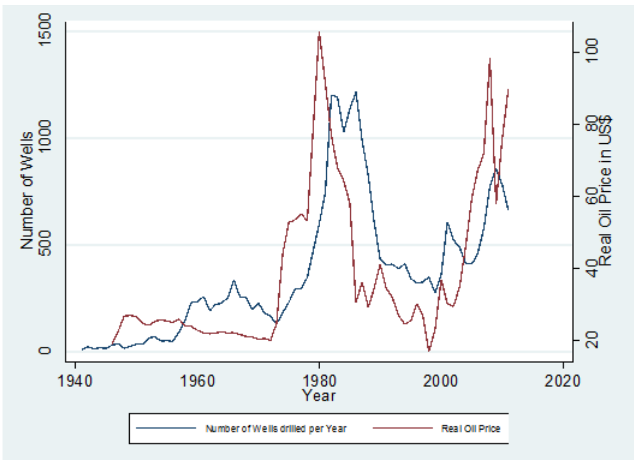
The evolution of knowledge of how to identify potentially oil-bearing formations comprises both advances in the theory of petroleum-bearing formations and ever-improving technology. In the early days of oil exploration, conspicuous targets were searched for oil, either without applying geology theory (e.g., surface pools in the form of natural oil seeps) or by using knowledge of geology (e.g., anticlines and salt domes). Investigation of the surface (topography) of a region could point to conspicuous areas of oil-bearing formations. In the 1920s and 1930s, aerial photography expanded the possibilities for mapping areas suitable for drilling. In the mid 1900s, seismic technology improved subsurface mapping for the location of potential petroleum-bearing formations. By and large, seismic activity produces sound waves that can be used to characterize subsurface formations, that is,

[52] There are other forms of well stimulation such as hydraulic fracturing.

sound waves are generated and recorded by receivers, and that information could be used to infer rock formations. The idea is to map the subsurface rock layers by using sound waves as different rock layers have different acoustical properties. The recorded sounds are processed and assembled for interpretation. Existing seismic and well information highlights the potential for exploration of large hydrocarbon resources. Computerization of seismic data provided a great leap forward for the extraction industry: A large body of data can be processed at high speed and precision. Another revolution in the oil industry was the $3D$ visualization that made possible a more reliable selection of the best targets to be drilled. Moreover, $4D$ visualization (repeated $3D$ through time) helped in the planning of well life-time operation. In the past decade, automated drilling (the evolution of automation in drilling) has been an ongoing topic of discussion. Modern technology helps in making decisions regarding the best drill sites. Computers and satellite images have improved the assessment of deposits. Nevertheless, ultimately it is only by drilling that a company can be certain that hydrocarbon deposits really exist. In other words, even with the use of modern technology, it is only by drilling that the existence of oil can be confirmed.

One important factor in the identification strategy employed in this paper is the relationship between drilling effort and international oil prices. Figure B.1 shows a striking correlation between wells drilled and international oil prices, in agreement with the literature on the role of oil prices (e.g., Mohn and Osmundsen (2008)).

Fig. B.1: Drilling in Brazil and International Oil Prices

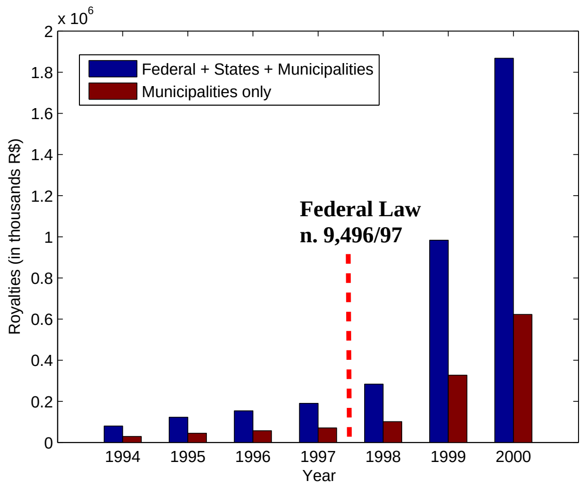


Notes. The figure shows the evolution of approximately 28,000 wells drilled (the universe of wells drilled in Brazil during the period from 1940 to 2013) and the evolution of the international price of crude oil during that period, defined as the monthly WTI (West Texas Intermediate) oil price in real US dollars per barrel.

Up to this point, we have described some general aspects of the upstream industry. The downstream industry includes the refineries, petrochemical plants, and distribution facilities (e.g., ports and terminals). Crude oil and natural gas are of little use in their raw state (Petroleum Extension Service (1997)). Refining and processing for the selection of certain groups of components (called "fractions") is what creates value. Refining means applying chemical processes to convert fractions into commercial products. Oil and gas vary in their hydrocarbon compounds and impurities (such as sulfur and metals). For instance, there are light crude oils as well as heavy and thick crude oils. The complexity of the composition of petroleum fractions leads to more than 2,000 individual refinery products (Fahim, Al-Sahhaf, and Elkilani (2009)). Examples of refined oil products are gasoline, jet fuel, kerosene, diesel fuel, and feedstocks for the petrochemical industry.

B.2 Royalties and Oil in Brazil

Fig. B.2: Distribution of Royalties: 1994–2000



Notes. In 1997, Federal Law n. 9,496/97 changed the rules for distribution of royalties.

The distribution of royalties from oil production in Brazil began in 1953. Federal Law n. 2,004/53 stipulated that 5% of the revenue from onshore oil production was to be distributed to states (80%) and municipalities (20%) in the form of royalties. Offshore oil royalties paid to states and municipalities were introduced by 1986. In 1997, Federal Law n. 9,496/97 changed the formula for the distribution of royalties (e.g., the international price of oil was used in the distribution formula for the first time). This led to a huge increase in royalty payments, as illustrated in Figure B.2, transforming it from a minor to a very significant source of income for municipalities.

The 1997 law requires an oil company to allocate between 5% and 10% of the value of the gross output in the form of royalties. The royalties are then divided among the three administrative levels in Brazil (federal government, states, and municipalities). A municipality is eligible to receive royalties based on (i) geography (if the production takes place within its territory or, in the case of offshore production, if it is a "facing" municipality, that is, there is an oil field that lies inside the municipality's maritime borders), (ii) oil-related infrastructure (if within its borders there is storage, transportation, or landing of oil and gas), and (iii) an equalization rule (there is a "special fund" that allocates part of the revenue from royalties to all the Brazilian municipalities). For some municipalities, royalties represent a significant part of their total revenue (more than half in extreme cases). According to ANP (Brazil's oil and gas industry regulator), over R$4.5 billion (circa US$2.2 billion) in oil windfalls was distributed to the Brazilian municipalities in 2010, which represented on average 2.5% of the total revenue of municipalities receiving oil windfalls.

For a much more detailed description of the history and technicalities of royalty payments in Brazil, see Caselli and Michaels (2013) and Monteiro and Ferraz (2012).

C Appendix: Treatment Intensity

We endeavor to measure the effect of treatment intensity. We ask how the outcome is related to the "dose" of the treatment. The literature on treatment intensity emphasizes the estimation of a weighting function to capture which group or observation is contributing the most to the results (e.g., Angrist and Imbens (1995), Frölich and Lechner (2010)). In the spirit of Angrist and Imbens (1995), our goal is to estimate a coefficient that can be interpreted as a weighted average of per-unit treatment effect. We thus estimate the following equation:

$$Y_{it} = \alpha + \tau prod_{it} + \beta_t' X_i + \gamma_i + \rho_t + \epsilon_{it}, \quad (4)$$

where we instrument the number of production wells ($prod_{it}$) with the number of discovery wells (field, subfield, and field extension wells) (($disc_{it}$)).[53] As an alternative measure of treatment intensity, we use the number of injection wells. The pressure in the reservoir is a key element in oil production, because that is the mechanism that drives oil and gas out of the reservoir. Normally, after some time, the pressure decreases and the oil company needs to (artificially) increase the pressure in the well. At that point, the oil company starts to drill "injection wells" to inject water, gas, chemicals, or steam to adjust for the falling pressure. Injection wells give us indirect information on the producing life of an oil field, because injection wells are used only to enhance production. Efforts to enhance production are costly and are dependent upon the potential oil recovery volume. In other words, it is viable to design injection wells to enhance production, but only if production is expected to be above a certain level. Therefore, we use injection wells as a measure of treatment intensity.[54] Note that while the t-statistic for the number of discovery wells in the first stage is always very high, the F-statistic for the GDP regressions are not particularly strong, indicating a potential weak instrument problem.

The signs of the various regressions are as before, and so we focus on quantifying the average per-unit effect on GDP *per capita* and urbanization. The results are reported in Table C.1. GDP *per capita* increases by 0.066% per production well and by roughly 1% per injection well. The urbanization rate increases by 0.007% per production well and by 0.15% per injection well. The coefficients for production wells are quite small. With the average producer MCA having 160 production wells, this gives an average impact of oil production of 160*0.0007=11.2%<20%. On the other hand, the coefficients for injection wells seem very large. This is a consequence of their ability to isolate the large production fields very well. In fact, only a handful of large fields onshore in the northeast or off the coast of Rio de Janeiro have any significant number of them. Our interpretation of these results is that large discoveries have a disproportionately large impact.

[53] We obtained production data by oil field from ANP for the year 2000 to construct production volume by MCA and compare it to the number of production wells. While the correlation between the two is high, it is higher for onshore than offshore production.

[54] Tabulations from Brazil support this choice. In the year 2000, for onshore fields, the MCAs with discovery wells and injection wells had much higher production volumes of both oil and gas than those that had discovery wells but no injection wells. In other words, the MCAs with injection wells were the ones with a lot of production.

Table C.1: Treatment Intensity

	Matched Dry Drilling			
	(1)	(2)	(3)	(4)
VARIABLES	ln GDP *per capita*	Urbanization Rate	ln GDP *per capita*	Urbanization Rate
Number of Production Wells	0.000664**	7.55e-05**		
	(0.000317)	(3.70e-05)		
Number of Injection Wells			0.0123**	0.00146*
			(0.00573)	(0.000871)
First Stage F-Stat.	6.98	15.92	6.29	31.21
MCA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	768	1,024	768	1,024
Number of MCAs	128	128	128	128
Geographical Controls	Yes	Yes	Yes	Yes
Initial Conditions	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE

Notes. Standard errors clustered at the MCA level. The main explanatory variable is the number of injection wells. The numbers of injection and production wells are instrumented with the number of discovery wells. Geographical controls and initial conditions have time-varying coefficients.
*** p<0.01, ** p<0.05, * p<0.1

NOT FOR PUBLICATION

D Appendix: Using Microdata from RAIS and Population Census

In Section 5.3, we use two sources of data. First, we use a matched worker–firm microdata from the Ministry of Labor's RAIS (Relação Anual de Informações Sociais). The RAIS dataset has information on each formal worker at each plant in Brazil. In 2000, there were 36,907,953 formal workers in the dataset. This information was useful to construct measures of average wages, as well as the numbers of workers and firms by skill and sector at the municipal level. We also calculate firm density and worker density, which are specified as the number of firms and workers, respectively, per square kilometer. We complement the analysis by using microdata from the 2000 population census, collected by the Brazilian Bureau of Statistics. The population census data allow us to calculate the fraction of workers employed in the formal sector by municipality, labor force participation, and sectoral employment shares.

In the analysis, we use cross-sectional data for the year 2000, because this is the first year for which high-quality data from both the employment and population censuses are available. The RAIS data have been collected annually since the late 1980s but are considered to have been of high quality only since the mid 1990s. The population census data are collected once per decade, making 2000 the first year in which it overlapped with reliable RAIS data.