Winning the oil lottery: the impact of natural resource extraction on growth

Tiago Cavalcanti[1,2] · Daniel Da Mata[2,3] · Frederik Toscani[4]

Abstract

This paper provides evidence of the causal impact of oil discoveries on local development. Novel data covering the universe of oil wells drilled in Brazil allow 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 local production and have positive spillovers. Workers relocate from informal, low productivity agriculture to higher value-added activities in formal services, increasing urbanization. The results are consistent with greater local demand for non-tradable services driven by highly paid oil workers.

Keywords Oil and gas · Economic growth · Local economic development · Structural transformation

Mathematics Subject Classification O13 · O40

✉ Tiago Cavalcanti
  tvdvc2@cam.ac.uk

  Daniel Da Mata
  daniel.damata@fgv.br

  Frederik Toscani
  ftoscani@imf.org

[1] Faculty of Economics, University of Cambridge, Cambridge, UK

[2] Sao Paulo School of Economics - FGV, São Paulo, Brazil

[3] Institute for Applied Economic Research (IPEA), Brasília, Brazil

[4] International Monetary Fund (IMF), Washington, USA

1 Introduction

There is a long tradition in economics of studying the impact of natural resource abundance on development, but no clear consensus has emerged in the literature. Nominal exchange rate appreciation and rent seeking can have adverse effects, as can volatility of revenues, but the large fiscal windfall associated with resource revenue can also foster development. Even when we abstract from nominal exchange rate movements and the impact of oil rents, the pure effect of the physical presence of a natural resource sector might drive up local prices—and therefore crowd out the development of other economic activities, bringing about negative effects on growth. On the other hand, the natural resource sector might also 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 "randomization", 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 focus on the direct impact of oil extraction rather than 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 oil production as the treatment, and its 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]

The baseline results show that locations in which oil was discovered had a roughly 30% 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 *per capita* GDP but no impact on agricultural GDP. While the measure of manufacturing GDP includes

---

[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.

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

[4] We discuss the endogeneity of oil production in greater detail in Sections 3 and 4.

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. Using historical data on employment shares by sector, we corroborate the GDP results by showing that the fraction of workers in the services sector increases significantly following oil discoveries. Additionally, we find evidence for an increase in urbanization of about 4% points. This increase in urbanization is consistent with the increase in services we document. We do not find any effect on population density or total worker density.

Distinguishing between onshore and offshore discoveries, we find that the results are entirely driven by onshore discoveries. We hypothesize that is because only onshore production (but not offshore) causes a local demand shock associated with the physical presence of the oil company and well paid oil workers. We find no detectable spillovers to neighboring municipality which we explain (jointly with the lack of impact on population density) by low inter-regional labor mobility.

In order to shed more light on the results, 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. Informality falls as a consequence of oil discoveries. The move from informal, low productivity rural work to the formal services sector explains the observed increase in urbanization and services GDP *per capita*. Lastly, the density of non-oil manufacturing firms and workers is not affected by oil discoveries.

The initial conditions of the local economies we study (large subsistence agriculture sector with very low productivity, small or non-existent manufacturing base) are likely to be crucial for the results we obtain. The impact of oil on manufacturing depends on the scale and specialization of the sector as well as the interplay between agglomeration effects and the crowding out effect from the resource boom (Allcott and Keniston 2018). In our setting of a developing country with little non-oil manufacturing in the affected areas, the presence of oil had no effect on manufacturing, but a strong effect on urban services and precipitated a decrease in the highly unproductive subsistence agriculture sector. Out of the large theoretical literature which tries to explain how natural resource abundance might affect economic outcomes (e.g., Corden and Neary 1982 and Krugman 1987), the framework proposed by Gollin et al. (2016) is most closely related to our results. In line with what we find, they illustrate how natural resource production can lead to urbanization as labor moves from rural food production to urban non-tradables.

Our results are robust to a variety of control groups, different control variables, and different sample periods. 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.

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 (Sachs and Warner 2001). However, cross-country evidence is sensitive to changing periods, sample sizes, and covariates (for an overview of the literature, see van der Ploeg 2011). Cotet and Tsui (2013b) for instance exploit cross-country variation in the size of oil endowments to show that oil does not hinder economic growth and is positively associated with health improvements.

One important strand of 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 this emergent literature are, among others, Michaels (2011), Monteiro and Ferraz (2012), Caselli and Michaels (2013), and Allcott and Keniston (2018). Caselli and Michaels (2013) study the effects of oil windfalls in offshore oil producing municipalities in Brazil and find little improvement in the provision of public goods or the population's living standards.[5] Our results complement theirs given that they (i) focus on the period when royalties became an important revenue for local governments in Brazil while we focus on the period before that distribution of royalties and (ii) they look at offshore production only while we find the positive demand shock impact is driven by onshore discoveries.

The main empirical challenge is to deal with the endogeneity of natural resource extraction, since many unobservable factors which affect economic development might be correlated with oil production and oil discoveries. Cust and Harding (2014), for example, show the important role institutions have in influencing the location of exploratory oil drilling. Since we exploit the randomness of oil discoveries conditional on exploration, Cotet and Tsui (2013a) is the closest in spirit to our identification assumption. In a cross-country sample of oil-producing countries, they exploit the randomness in the size of discoveries to investigate the impact of oil reserves on conflicts.

Our paper stands out from the existing literature in at least three important respects: Firstly, our 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 an oil-producing country, while attention has mostly been limited 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 a developing country.[6]

It is important to stress 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.

This article proceeds as follows. Section 2 describes the data we use, while Sect. 3 presents the empirical model. In Sect. 4 we describe the quasi-experiment which we exploit in detail, including a description of the institutional environment in Brazil and a short background on the technicalities of drilling for oil. Sect. 5 presents the results. Sect. 6 concludes.

2 Data

In this section, we describe the data used to study the impact of oil on economic development at the municipal level in Brazil. Our period of study is from 1940 to 2000, starting just before the first successful oil discovery in 1941. One complication when dealing with municipalities in Brazil is the process of detachments and splits that have taken place over the years. In 1940 there were 1,574 municipalities, while in 2000 there were 5,507. In order to deal with the

---

[5] Monteiro and Ferraz (2012) use windfalls in Brazil to study local political and economic outcomes. See also Brollo et al. (2013) for an analysis of fiscal windfalls in Brazil. See Acemoglu et al. (2013), Aragón and Rud (2013), and Dube and Vargas (2013) for studies on the local effects of resource wealth.

[6] Our paper is also related to the literature on agglomeration externalities, especially the branch that investigates the impact of interventions on the concentration of economic activity (e.g., Davis and Weinstein 2002; Greenstone et al. 2010). Lastly, our focus on sectoral GDP and employment connects us to studies on the determinants of structural transformation, particularly the ones focusing on the role of the oil sector (Kuralbayeva and Stefanski 2013; Stefanski 2014).

Table 1 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	1067	2556	3623
Development wells	Producer	1368	9101	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	1017	1090
Other	Abandoned	421	554	975
	Special	62	369	431
	Missing category	30	171	201
	Total	4031	16,021	20,052

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
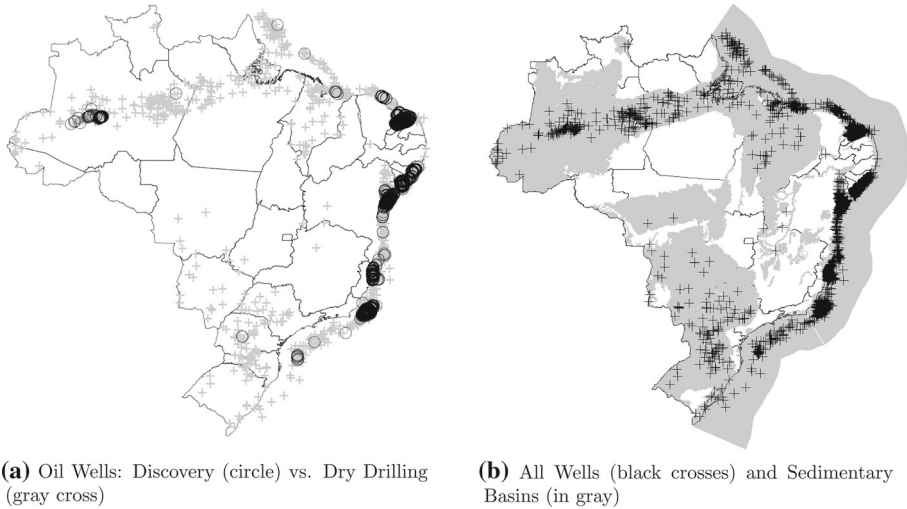
detachments, we use the concept of a Minimum Comparable Area (MCA), which consist of sets of municipalities whose borders were constant over the study period. We thus aggregate municipalities to 1,275 MCAs.[7]

Oil discoveries and oil production: 1940–2000 To obtain information on which municipalities discovered oil and are producing oil we use a well level dataset from *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis* (ANP), the Brazilian oil and gas industry regulator. The dataset contains detailed information on the universe of wells drilled in Brazil: 20,052 wells 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, among others). Furthermore, we have information on the viability of exploring the oil deposit (when oil was found) and on whether the oil company started production by drilling production wells.

Table 1 shows the number of wells by category. 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).[8] However, wells can be broadly classified as exploratory wells and development wells. Exploratory wells are drilled to test for the presence of oil, while wells drilled inside the

---

[8] We obtained more the 50 different classifications from the dataset, but we were able to aggregate all of them into few major categories (see Table 1). One limitation of the dataset is that information on the amount of oil produced by each individual producer well is available only from the year 2000 onward. See Appendix B.1 for a detailed explanation of the types of wells.

(a) Oil Wells: Discovery (circle) vs. Dry Drilling (gray cross)

(b) All Wells (black crosses) and Sedimentary Basins (in gray)

Fig. 1 Oil wells in Brazil: 1940–2000. *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 **a**, wells with Oil Discovery are in circles, Dry Wells are in the format of a cross. **b** Shows the locations of sedimentary basins in Brazil (in gray) and the universe of oil wells (black crosses). 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.)

known extent of the field are called development wells (e.g., producer wells). Unsuccessful drilling is classified as a dry hole in both exploratory and development categories. Figure 1 shows the geographic distribution of drilling and discoveries in Brazil and highlights that oil drilling in Brazil is concentrated in sedimentary basins (which is where oil can potentially be found—see Sect. 4).

To match wells and MCAs we proceed 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.[9]

Local economic development: 1940–2000 We combine data from several sources to obtain as much information as possible on measures of local economic development in Brazil. In order to construct historical outcomes at the municipal level, we use two main data sources: Population Censuses and Economic Censuses.

The Population Censuses provide us with a reliable long-running source of information on population characteristics at the municipal level. From the Population Censuses (of 1940, 1950, 1960, 1970, 1980, 1991, 1996, and 2000), we obtained data on population counts, population density, urbanization rate, education, and employment.[10] We group sectoral employment categories so that they are consistent over time. We obtain the following six categories: (i) agriculture and fishing, (ii) manufacturing including extractive activities, (iii) retail, (iv) transportation, (v) public sector and (vi) services.[11]

---

[9] As a robustness check, we also use an alternative method to allocate offshore wells to MCAs (see Sect. 5.2).

[10] Densities are specified as population per square kilometer. The urbanization rate is the proportion of the population living in urban areas. Education data is from 1970 to 2000.

[11] For time consistency, retail includes activities related to the selling of a broad range of goods to consumers, while services include a wide range of activities related to health, education, recreation, sports, hotels, restaurants, financial, personal care and other services.

Gross Domestic Product (GDP) data are from Economic Censuses (of 1949, 1959, 1970, 1975, 1980, and 1985), sector surveys in 1996, and from the national accounts of 2000. Reis et al. (2004) constructed municipal-level GDP from historical Economic Censuses, from where it is possible to calculate GDP through the production approach. Since 1949, the Censuses provide data on the value of total outputs and total costs (a proxy for intermediate goods) at the municipal level. More precisely, the Censuses provide total output and total input by economic sector so it is possible to construct value-added figures by sector. Sectoral GDPs were then added so as to calculate the total municipal GDP. Oil and Mining were included in the manufacturing GDP. The municipal GDP is deflated using the national implicit price deflator.[12]

Microdata To improve on the analysis, we use cross-sectional microdata for the year 2000. We use a matched worker–firm dataset from the Ministry of Labor's RAIS (Relação Anual de Informações Sociais). The RAIS data have been collected annually since the late 1970s but are considered to be of high quality only since the mid 1990s. Since the population census data are collected once per decade, 2000 is the first year in which reliable RAIS data overlapped with a population census. 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. We use this information 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. Since RAIS only covers workers in the formal sector, we complement it with microdata from the 2000 Population Census, which allow us to obtain the fraction of workers employed in the formal sector.

Geography Data on average temperature, average rainfall, and average altitude come from Ipeadata.[13] Further data comprise the latitude and longitude of each MCA, distance to the closest state capital, as well as geographical indicators of its location (on the coast, in the Amazon region, and in the semiarid region).[14]

2.1 Stylized fact: a first look at the data

Figure 2a, b show GDP *per capita* for the period 1940–2000 in the states of Rio de Janeiro and Sergipe (two important oil-producing states in Brazil), 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, 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 increased 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 which

---

[12] Appendix A.1 details the calculation of municipal GDP, and the caveats one should bear in mind when using the GDP data. In the empirical exercises, we also use sectoral employment data concomitantly to the sectoral GDP.

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

[14] 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.

(a) GDP *per capita* in Rio: 1940–2000   **(b)** GDP *per capita* in Sergipe: 1940–2000
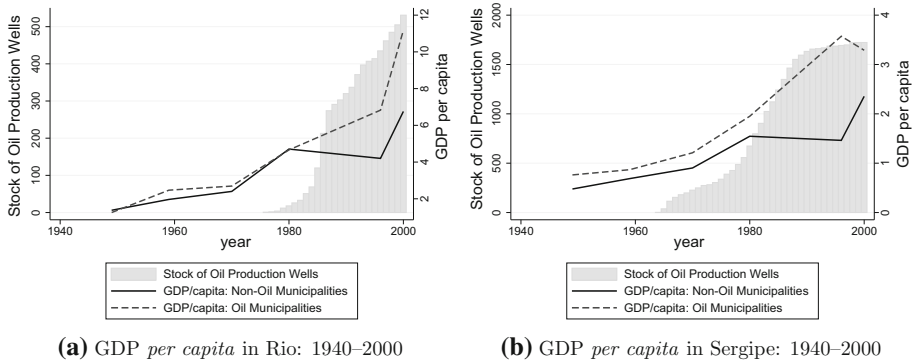
Fig. 2 GDP *per capita* in oil and non-oil municipalities. *Notes* Figure shows *per capita* GDP in municipalities of the states of **a** Rio de Janeiro and **b** Sergipe in which oil was discovered during the period 1940 to 2000 (dotted line) and those in which it was not (solid line). Rio de Janeiro 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 Sergipe was discovered in the mid 1960s

non-oil sectors are affected and whether the spillovers of oil production to other sectors are positive or negative.

3 The empirical model

We aim to recover the impact of oil on economic development at the local level. We present the basic regression model in this section, and discuss the quasi-experiment we exploit to obtain a causal coefficient in Sect. 4. Let Y_{it} be a measure of economic development in MCA i and year t. Our empirical model is the following Difference-in-Difference specification:

$$Y_{it} = \alpha + \tau T_{it} + \beta_t' X_i + \gamma_i + \rho_t + \epsilon_{it}, \qquad (1)$$

where X_i are time-invariant MCA characteristics, including the pre-treatment level of the dependent variable, ϵ_{it} is an error term, ρ_t denotes year fixed effects and γ_i denotes MCA fixed effects. The source of cross-sectional and time variation is given by an indicator for the *production of oil* T_{it}. The coefficient of interest is τ. Time fixed effects control for shocks common to all Brazilian MCAs, while the MCA fixed effects capture time-invariant MCA characteristics such as location, geology, or distance to the coast. Some of those time-invariant geographic characteristics might have time-varying effects (the importance of being located on the Coast might have increased as Brazil was integrated more closely into the world economy, for example). To address this we explicitly include longitude, latitude, an indicator for being located in the Amazon, and an indicator for being located on the coast in the vector of controls X_i and allow for time-varying coefficients β_t. We also include measures of economic development in 1940 (before the first successful oil discovery) in X_i to capture initial conditions. When the dependent variable is expressed as a logarithm, the percentage difference between oil and non-oil municipalities is calculated from the estimated τ as $100 * [e^\tau - 1]$. 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.

4 The oil lottery: a quasi-experiment

An estimation of Eq. (1) above would expose our results to a major concern as oil production is likely to be endogenous to local economic conditions. For example, production might be more attractive close to large urban centers, might be influenced by strategic behavior regarding production quotas, or might occur in some regions but not others because of political economy considerations. The endogeneity related to production might even be more problematic because the transportation of oil (and gas) requires a substantial investment in infrastructure such as pipelines.

A first step to address the endogeneity issue is to use discoveries instead of production as the explanatory variable. Discoveries are arguably more exogenous than production. This might not go a long way in overcoming endogeneity concerns, however. Recent papers such as Cust and Harding (2014) show that institutions are an important driver of oil exploration and discoveries. We therefore need to take a further step back to try and identify exogenous variation in discoveries. We obtain this exogenous variation by exploiting the quasi-experiment generated by the randomness in the success of exploratory oil drilling. In other words, our key identifying assumption is that conditional on exploratory drilling taking place, a discovery is unrelated to local economic conditions.

In practice, using the well data, we first restrict our analysis to municipalities with drilling and then construct an indicator for whether a discovery was made and another for whether oil is produced. The dummy for production (T) follows immediately from the well data — it 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 (Z) 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 meaningful discovery includes a field or subfield discovery and subsequent field extension discoveries to delineate the size of the oil field (see Appendix B.1).

We thus obtain the following numbers regarding drilling and oil discoveries:

- Total number of MCA units = 1275
- Drilling MCAs = 222
- All discoveries MCAs = 64
- True discoveries MCAs = 45

We introduce the quasi-experiment in four steps. First, we briefly present the econometric model. Second, we discuss the institutional setting in Brazil. Third, we provide background on oil drilling to justify the identifying assumption qualitatively. And fourth and last, we conduct formal tests to lend support to the identifying assumption.

4.1 Econometric framework

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 ITT $= \mathbb{E}[y_i|Z_i = 1] - \mathbb{E}[y_i|Z_i = 0]$.
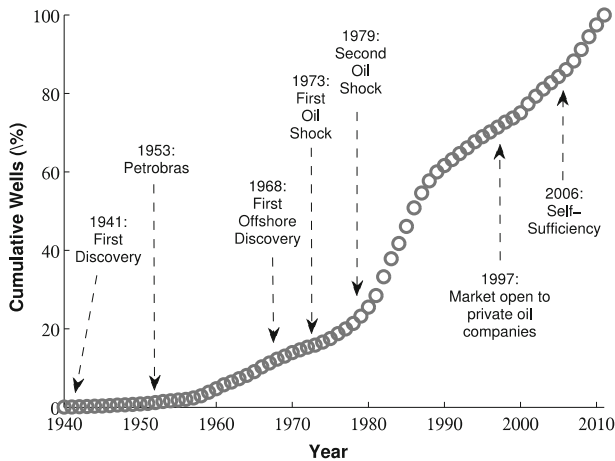
Fig. 3 Events and oil drilling: 1940–2011. *Notes* Figure shows the cumulative percentage of oil wells drilled in Brazil during the period from 1940 to 2011

The oil discovery dummy is represented by Z_{it} (our 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. Following Eq. 1 we assume an additive and linear empirical specification to estimate an ITT effect, as follows:

$$Y_{it} = \alpha + \tau_{ITT} Z_{it} + \beta'_t X_i + \gamma_i + \rho_t + \epsilon_{it}. \tag{2}$$

We will also use discoveries to instrument for production to recover a coefficient which can be interpreted as a Local Average Treatment Effect (LATE).

4.2 The institutional setting in Brazil

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 viable onshore well was drilled. The first oil discovery from an offshore well took place in 1968. Figure 3 summarizes domestic and international events related to oil exploration and production in Brazil.[15]

During most of our period of interest, 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.[16] From 1953 to 1997, only one company was allowed to drill for oil in Brazil: the government-controlled Petrobras.[17]

---

[15] In 2011, Brazil was the world's 13th largest producer of oil and gas, with 2.2 million 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).

[16] 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.

[17] 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

Petrobras is an integrated exploration and production company whose activities encompass all phases of the oil supply chain.

Royalties did not play an important role for local government finances for most of our sample period. Only in 1997 a change in the allocation rule led to huge increases in royalty payments to municipalities and turned them into a key source of local government revenue. Prior to the reform, royalties accounted for roughly 3% of municipal budgets in oil producing municipalities. This allows us to claim that in our analysis we will identify mainly the direct impact of oil production on local economic development rather than the indirect impact via a fiscal windfall (see online Appendix B.2 for more details on royalties in Brazil).

4.3 The success of oil drilling as a randomization

Oil and gas exploration is known to be 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.

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 et al. 1995). 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."[18] Our data also support the idea that discovering oil can be viewed as a "lottery" (where drilling for oil is akin to buying the lottery ticket): For every exploration well drilled which was successful, there were many more unsuccessful ones—a ratio of roughly one over four (recall Table 1 in Sect. 2).

4.4 The identifying assumption in practice

In this part, we provide evidence on the exogeneity of drilling success. We then test—as we work with a difference-in-difference design—for parallel trends and the stable unit treatment value assumption (SUTVA) between the treatment and control groups. Lastly, we discuss the covariate overlap between the treatment and control group.

Footnote 17 continued

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.

[18] According to Harbaugh et al. (1995), "luck is obviously a major factor in exploration." Fustier et al. (2016) estimate that exploration success rates worldwide fluctuated around 20% between the period from 1960 to 2010.

Table 2 Discoveries, conditional on drilling

	(1)	(2)	(3)
Dependent variable	Number of discovery wells		Drilling success ratio
	OLS	Poisson	Linear probability
Urbanization in 1940	0.741	− 0.814	0.120
	(9.777)	(1.562)	(0.125)
ln Pop. density in 1940	2.771	0.416	− 0.0145
	(2.499)	(0.347)	(0.0247)
ln GDP *per capita* in 1949	2.959	0.548	− 0.00265
	(2.239)	(0.344)	(0.0288)
ln Worker density in 1940	−2.436	− 0.298	0.0129
	(2.857)	(0.384)	(0.0247)
Semiarid indicator	10.80	1.389 ∗ ∗	0.102
	(7.663)	(0.545)	(0.0689)
Amazon indicator	2.761	− 0.335	− 0.0274
	(4.005)	(0.814)	(0.0642)
Coastal indicator	10.84 ∗ ∗	1.732∗∗∗	0.0596
	(5.456)	(0.667)	(0.0425)
Distance to closest state capital (km)	0.000687	0.000501	0.0000045
	(0.0105)	(0.00225)	(0.000152)
Constant	− 1.853	0.358	0.0731
	(4.300)	(0.837)	(0.0464)
Observations	222	222	210

This table reports the correlation between oil discoveries and economic and geographical characteristics. Robust standard errors in parentheses. The cross-section regressions are for the 222 Minimum Comparable Areas (MCAs) in which Petrobras drilled for oil. The dependent variables are the total number of discovery wells in 1940–2000 and the drilling success ratio in 1940–2000. 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, worker 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. We also control for the distance to closest state capital (in km).
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Exogeneity of discoveries and parallel trends Table 2 suggests that the success of oil drilling is exogenous to local economic conditions in 1940 (pre-oil exploration in Brazil). We regress (i) the number of exploratory wells with a discovery between 1940–2000 and (ii) the ratio of successful drilling to unsuccessful drilling in the same period on pre-treatment characteristics. We find that both the number of discoveries and the drilling success ratio are unrelated to pre-treatment local economic characteristics. Table 2 shows that the number of discoveries turns out to be correlated with certain geographical characteristics; we return to this point below. Reassuringly, the success ratio is uncorrelated with all controls (Column (3)). This supports the notion that, conditional on drilling taking place, success seems to be a lottery.

We also check whether conditional on a first discovery, additional drilling attempts (and thus discoveries) are unrelated to local economic development (see Table A.3 in online Appendix A). 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,

Table 3 Parallel trends

Variables	Parallel trends			
	(1)	(2)	(3)	(4)
	ln GDP *per capita*	ln population density	ln workers density	Urbanization rate
Discovery (next decade)	− 0.0782	− 0.00297	0.0108	− 0.00349
	(0.0768)	(0.0393)	(0.0525)	(0.0148)
Dummy for discovery (past decade)	Yes	Yes	Yes	Yes
Dummy for discovery (current decade)	Yes	Yes	Yes	Yes
MCA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1332	1776	1330	1776
Number of MCAs	222	222	222	222
Geographical controls	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE

Standard errors clustered at the MCA level. This table shows the results of estimating Eq. 3. 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, worker density in 1940, and population density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

this could bias our results (in particular, when using the "True Discovery" dummy). Unsurprisingly, drilling attempts increase significantly after an initial discovery was made. After a first discovery, naturally Petrobras intensifies its efforts in that particular area. Importantly, however, there is no indication that drilling increases *more* in MCAs with higher *per capita* GDP, higher urbanization or with a higher population or employment density.

We test for parallel trends by estimating

$$Y_{it} = \alpha + \tau_{t-1} Z_{i,t-1} + \tau_t Z_{i,t} + \tau_{t+1} Z_{i,t+1} + \beta'_t X_i + \gamma_i + \rho_t + \epsilon_{it}, \tag{3}$$

where $Z_{i,t}$ is equal to one only in the decade in which a discovery takes place and $Z_{i,t-1}$ and $Z_{i,t+1}$ indicate a discovery in the previous period and the next period, respectively. For the parallel trend assumption to hold we require $\tau_{t+1} = 0$, i.e., the coefficient on future discoveries should be zero.[19] Table 3 shows that indeed the coefficient on the lead is zero, implying that there are no differences in current outcomes regarding *per capita* GDP, population density, worker density, and urbanization rate between municipalities which will discover oil and those that will not.[20] The no-anticipatory effect reinforces the notion that discoveries are unrelated to pre-treatment economic conditions.

The stable unit treatment value assumption Spillovers from treated municipalities to non-treated ones pose another potential threat to our identification. To test for possible SUTVA violations, we implement a placebo treatment where neighbors of discovery municipalities (those most likely to be affected by spillovers) are defined as the treated units. We use all

---

[19] We did not include additional leads to test the parallel trends assumption over more than one period because of the limited number of time periods in our panel.

[20] We will return to discussing the coefficient on the lag, which provides valuable information on the dynamic impact of a discovery, in the results' section.

other Brazilian municipalities (excluding those who did have the actual discovery) as the control group. As Table A.4 in online Appendix A shows, there is no detectable impact on municipalities whose neighbors discovered oil. We can thus not find any evidence of spillovers outside of the discovery municipality. We return to the topic of a lack of spillovers when we discuss the baseline results in Sect. 5.

The overlap between the assigned to treatment and control groups As Table 2 showed, the number of discoveries is correlated with some geographical characteristics. Ex-post, discoveries tended to be disproportionately located on the coast, while locations with exploratory drilling and no discoveries are more likely to be land-locked. When the covariate overlap between the treatment and control groups is limited, results based on linear regression methods can be sensitive to changes in specification (Imbens and Wooldridge 2009). Therefore, we formally check for overlap using normalized (or standardized) differences (Rubin 2001 and Imbens and Wooldridge 2009).[21]

While the overlap is good for initial economic conditions, Table 4 indicates that the overlap between the assigned to treatment and control group is not ideal for certain geographical variables (consistent with Table 2). 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).[22] For this subsample, we choose the 64 municipalities with the highest propensity score and call this control group "matched dry drilling." Figure 4 shows maps with the locations of the two control groups we obtain. Figure 4a shows the places with discoveries and the set of all MCAs where drilling took place and no oil was found. Figure 4b displays the matched dry-hole subpopulation.

5 Results

5.1 Baseline results

Socio-economic variables Table 5 shows the baseline ITT results (Eq. 2) using the "All Discovery" dummy as our treatment assignment. We show results for both the dry drilling control group and the matched dry drilling sample. The key independent variable is the discovery dummy. The dependent variables are *per capita* GDP, population density and worker density, which are expressed as logs. An additional dependent variable is the urbanization rate, which is 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 13.3–15.7% over a 60-year period as a result of oil discoveries. Population density, worker density, and the urbanization rate are unaffected by oil discoveries in this specification.

As discussed earlier, the "All Discovery" dummy has some conceptual drawbacks. It is also a weaker predictor of oil production than the "True Discoveries" dummy. 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).[23]

---

[21] Standardized differences are not influenced by sample size, unlike *t* tests and other statistical tests. Imbens and Wooldridge (2009) suggest that the normalized difference should be below 0.25.

[22] In the propensity score analysis, we used the available data described in Sect. 2. See online Appendix A.2 for more details.

[23] Implicitly, other recent papers on the impacts of oil abundance have also defined relevant discoveries. For example, Michaels (2011) uses a threshold of 100 million barrels of reserves, and Allcott and Keniston (2018) use a cutoff in production of US$100 per inhabitant.

**Table 4** Overlap between assigned to treatment and control groups

Variable		(I) Oil discovery	(II) Dry drilling	(III) Matched dry drilling
Pop. density 1940	Mean	32.89	30.33	35.09
	SD	51.35	132.29	104.47
	Standardized difference	–	0.018	− 0.019
Urbanization 1940	Mean	0.27	0.22	0.24
	SD	0.18	0.18	0.2
	Standardized difference	–	0.207	0.111
GDP per capita 1949	Mean	0.67	0.88	0.69
	SD	0.42	0.89	0.75
	Standardized difference	–	− 0.219	− 0.027
Manufacturing/GDP 1949	Mean	0.19	0.13	0.13
	SD	0.15	0.16	0.14
	Standardized difference	–	0.258	0.265
Services/GDP 1949	Mean	0.38	0.37	0.4
	SD	0.2	0.21	0.23
	Standardized difference	–	0.022	− 0.072
Agriculture/GDP 1949	Mean	0.43	0.51	0.48
	SD	0.24	0.24	0.26
	Standardized difference	–	− 0.227	− 0.144
Worker Density 1940	Mean	10.83	9.75	10.8
	SD	16.78	43.35	29.65
	Standardized difference	–	0.023	0.001
Share of workers in services sector 1940	Mean	0.078	0.064	0.073
	SD	0.055	0.058	0.065
	Standardized Difference	–	0.166	0.046
Share of workers in public sector 1940	Mean	0.015	0.016	0.019
	SD	0.014	0.019	0.024
	Standardized difference	–	− 0.034	− 0.125
Avg. rainfall	Mean	118.46	127	122.23
	SD	38.79	43.65	51.44
	Standardized difference	–	− 0.146	− 0.059
Avg. temperature	Mean	24.95	23.96	24.35
	SD	1.9	2.97	2.7
	Standardized difference	–	0.281	0.182
Latitude	Mean	− 11.88	− 13.72	− 12.62
	SD	6.44	9.67	8.6
	Standardized difference	–	0.158	0.069
Longitude	Mean	− 40.65	− 46.94	− 43.5
	SD	6.46	7.31	7.6
	Standardized difference	–	0.645	0.286

Table 4 continued

Variable		(I) Oil discovery	(II) Dry drilling	(III) Matched dry drilling
Coastal indicator	Prop.	0.59	0.3	0.53
	Standardized difference	–	0.431	0.089
Semiarid indicator	Prop.	0.19	0.15	0.23
	Standardized difference	–	0.080	− 0.081
Amazon indicator	Prop.	0.08	0.3	0.17
	Standardized difference	–	− 0.413	− 0.202
	Number of MCAs	64	158	64

Oil discovery is the treated group of 64 MCAs. Two control groups are shown: 158 MCAs where drilling took place but nothing was found (Column II: "Dry Drilling"), and propensity score matched sample of 64 MCAs where drilling took place but nothing was found (Column III: "Matched Dry Drilling"). Imbens and Wooldridge (2009) suggest that the normalized difference should be below 0.25 (in absolute values). In the propensity score analysis, we used the available data described in Sect. 2. See online Appendix A.2 for more details on the propensity score analysis

(a) Discovery (Dark Gray), Dry Drilling (Light Gray)

(b) Discovery (Dark Gray) and Matched Dry Drilling Sample (Light Gray)
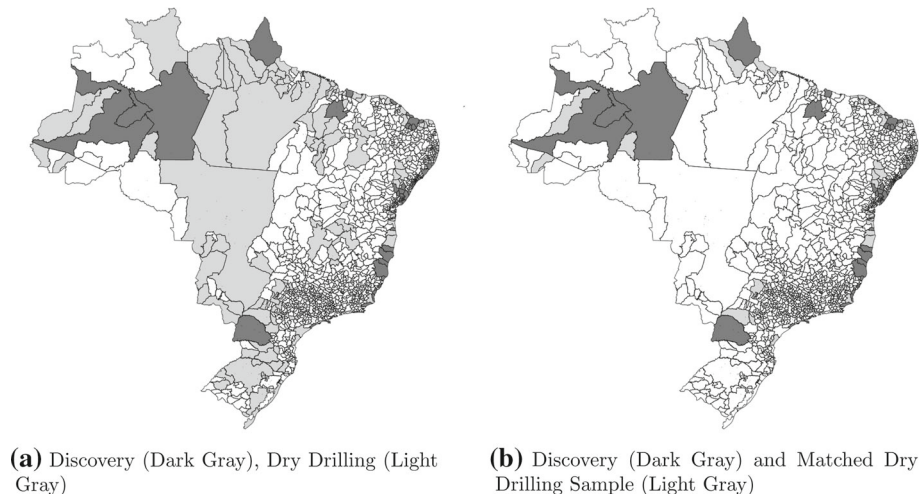
Fig. 4 Treatment and control groups. *Notes* Figures show 1275 Minimum Comparable Areas (MCAs) in 1940. The discovery dummy is the "All Discoveries" dummy (which is equal to one when at least one field, subfield, or field extension discovery was made in the MCA)

Table 6 shows the baseline ITT results using our preferred treatment assignment ("True Discoveries"). Unsurprisingly, the coefficients are markedly higher than in Table 5. The increase in *per capita* GDP is estimated at 27.9–29.6%. While population density and worker density are 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 meaningful 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.

Table 5 Intention-to-treat effect of all oil discoveries: socio-economic outcomes

Variables	Dry drilling				Matched dry drilling			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln GDP per capita	ln population density	ln workers density	Urbanization rate	ln GDP per capita	ln population density	ln workers density	Urbanization rate
Discovery dummy	0.125*	−0.0390	−0.0720	0.0283	0.146*	−0.0400	−0.0637	0.0253
	(0.0728)	(0.0579)	(0.0669)	(0.0187)	(0.0783)	(0.0626)	(0.0713)	(0.0199)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1332	1776	1330	1776	768	1024	767	1024
Number of MCAs	222	222	222	222	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	FE	FE	FE	FE	FE	FE	FE	FE

This table shows the results of estimating Eq. 2. 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, Worker Density in 1940, and Population Density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6 Intention-to-treat effect of true oil discoveries: socio-economic outcomes

Variables	Dry drilling				Matched dry drilling			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln GDP per capita	ln population density	ln workers density	Urbanization rate	ln GDP per capita	ln population density	ln workers density	Urbanization rate
Discovery dummy	0.246***	−0.00864	−0.0475	0.0443**	0.259***	−0.0127	−0.0394	0.0430**
	(0.0856)	(0.0676)	(0.0758)	(0.0202)	(0.0910)	(0.0731)	(0.0814)	(0.0213)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1332	1776	1330	1776	768	1024	767	1024
Number of MCAs	222	222	222	222	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	FE	FE	FE	FE	FE	FE	FE	FE

This table shows the results of estimating Eq. 2. 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, Worker Density in 1940, and Population Density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Sectoral GDP and sectoral employment shares To understand whether the increase in *per capita* GDP is purely mechanical, in the sense that there are no spillovers from oil production to other sectors of the economy, we investigate the impact of oil discoveries on sectoral GDP in Table 7.[24] 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 do not allow us to do so. As such, it is not surprising that manufacturing GDP increases significantly with oil discoveries. Importantly, however, services GDP increases by about 25%, while agricultural GDP is unaffected (the point estimate is negative but insignificant).

Table 7 (Columns (4)–(9)) also looks at how sectoral employment shares are affected by discoveries and finds results consistent with the estimated impact on sectoral GDP. We find that in oil municipalities an important structural transformation occurs: workers reallocate from the agricultural sector to the services sector. To a lesser degree, the share of workers in the public sector also increases. In terms of raw numbers, the results indicate a reduction in agricultural employment of roughly 3500 workers (out of total employment of 82,000 in 2000, 50,000 of which in agriculture) for the average MCA. Public sector and services employment increase on average by about 700 and 3000, respectively.

A plausible hypothesis is that local demand for non-tradables from well-paid oil workers and the oil company lead to an expansion of the services sector which attracts rural agricultural workers to move to the local urban agglomeration. At the same time additional local government revenues allow the municipal government to expand employment.

Onshore versus Offshore discoveries Distinguishing between onshore and offshore discoveries allows us to study this possibility in more detail. In particular, 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 off the coast of Rio de Janeiro, and most personnel associated with offshore production is stationed in only one location (the municipality of Macaé).

The disaggregated results for onshore and offshore discoveries are shown in Table 8 (in Columns (1)–(10)). For onshore discoveries we use municipalities with onshore drilling and no discoveries as the control group, while for offshore discoveries we use municipalities with offshore drilling and no discoveries as the control group. Municipalities with both onshore and offshore drilling are excluded from the analysis.

We find a large positive impact of onshore discoveries on local economic development but no impact of offshore discoveries. In fact, for offshore discoveries the coefficients are estimated to be equal to zero with some precision in Columns (6)–(10). For manufacturing GDP *per capita* the estimated coefficient is positive but not significantly. For onshore discoveries, services GDP *per capita* increases by 43%, the urbanization rate increases by over 8% points, while the fraction of services workers in the local economy increases by over 5% points. Additionally, the fraction of public sector workers increases by roughly 1% point. We interpret these results as support for the hypothesis that a structural shift towards the services sector is caused by a local demand shock in municipalities with onshore discoveries. Since this demand shock is absent after offshore discoveries there is no effect there. The impact on public sector employment might be due to two factors: first, there was a (very) small impact on government revenues even before 1997 due to royalties. In 1995, the first year for which we have data on royalties at the municipal level, they made up 2.84% of government revenue for onshore municipalities. Second, the increase in local activity generated additional local

---

Table 7 Intention-to-treat effect of oil discoveries: sectoral GDP and sectoral employment

Variables				Matched dry drilling					
					Sectoral employment: shares				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln agriculture GDP *per cap*	ln manufacturing GDP *per cap*	ln services GDP *per cap*	Agriculture	Manufacturing	Retail	Transportation	Public sector	Services
	Sectoral GDP								
Discovery dummy	0.0664	0.456**	0.215**	−0.0419*	−0.00768	0.00152	0.00266	0.00947***	0.0359***
	(0.109)	(0.189)	(0.104)	(0.0223)	(0.0151)	(0.00457)	(0.00313)	(0.00305)	(0.0114)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	765	764	765	767	767	767	767	767	767
Number of MCAs	128	128	128	128	128	128	128	128	128
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE	FE	FE	FE

Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as "True Discovery". The initial conditions with time-varying coefficients are: GDP *per capita* in 1949, Urbanization rate in 1940, Worker Density in 1940, and Population Density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8 Onshore versus offshore discoveries

Variables	Control group: dry drilling onshore					Control group: dry drilling offshore		
	ln GDP *per capita*		Urbanization rate	Employment share		ln GDP *per capita*		Urbanization rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Manufacturing	Services		Services	Public sector	Manufacturing	Services	
Onshore discovery	0.632**	0.359**	0.0866**	0.0585**	0.00909**			
	(0.272)	(0.117)	(0.0245)	(0.0131)	(0.00389)			
Offshore discovery						0.206	0.0488	– 0.00098
						(0.332)	(0.233)	(0.0487)
Onshore discovery * pre-1970								
Onshore discovery * post-1970								
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1007	1003	1352	1012	1012	322	324	432
Number of MCAs	169	169	169	169	169	54	54	54
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE	FE	FE

Table 8 continued

Variables	Control group: dry drilling offshore		Control group: dry drilling onshore				
	Employment share		ln GDP *per capita*		Urbanization rate	Employment share	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Services	Public sector	Manufacturing	Services		Services	Public sector
Onshore discovery							
Offshore discovery	0.027	0.00571					
	(0.0234)	(0.00435)					
Onshore discovery * pre-1970			0.684*	0.545***	0.0967***	0.0788***	0.0080*
			(0.395)	(0.156)	(0.0356)	(0.0171)	(0.0048)
Onshore discovery * post-1970			0.554	0.0842	0.0698***	0.0239**	0.011*
			(0.357)	(0.112)	(0.0220)	(0.0114)	(0.0061)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	320	320	1007	1003	1352	1012	1012
Number of MCAs	54	54	169	169	169	169	169
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE	FE

Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. The control group in Columns (1)–(5) and in Columns (11)–(15) is the group of municipalities with onshore drilling and no discoveries ("Dry Drilling Onshore"). The control group in Columns (6)–(10) is the group of municipalities with offshore drilling and no discoveries as the control group ("Dry Drilling Offshore"). The initial conditions with time-varying coefficients are: GDP *per capita* in 1949, Urbanization rate in 1940, Worker Density in 1940, and Population Density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

revenues via local tax collection. Property taxes, for example, are collected locally and could have benefited from the increased urbanization.[25]

Historical versus contemporaneous effects To gain additional insights, we split discoveries into pre- and post-1970. 1970 is a somewhat arbitrary cut-off based on the mid-point of our sample period. As virtually all offshore discoveries took place after 1970, this exercise essentially allows us to split the onshore discovery sample to explore the medium-run versus long-run effects of oil discoveries. It has the additional advantage that it allows us to verify whether the difference in results between onshore and offshore discoveries is simply due to different timing of the discoveries.

Columns (11)–(15) of Table 8 show how onshore discoveries before and after 1970 impact our relevant outcomes. Pre-1970 discoveries are associated with significantly larger coefficients. They have led to large increases in *per capita* GDP, urbanization, workers in the services sector, and workers in the public sector. Post 1970 onshore discoveries have a similarly sized point estimate on manufacturing GDP (as one would expect if production volumes are similar across the two groups) but the coefficient is imprecisely estimated. The results suggest that later discoveries increase urbanization and the share of workers in the services sector by only about 60% and 30%, respectively, as much as earlier discoveries. While power is a concern in these regressions, the results nevertheless offer some indicative evidence for long-run agglomeration effects associated with oil discoveries. Additionally, we can rule out that the lack of impact associated with offshore discoveries is purely a matter of the timing of the discoveries.

Growth effects So far we have focused on level effects to study long-run results. As an alternative, we estimate Eq. 2 in growth rates rather than levels (to facilitate quantitative interpretation, the growth rate of the dependent variables is annualized). The discovery dummy is set to one only in the decade of discovery (as in the specification to test for parallel trends (Eq. 3)). We include one lag of the discovery dummy in the regressions. As Table A.5 in online Appendix A shows, the growth effects are consistent with the previous analysis of the long-run level effect. Interestingly, it seems to take a number of years for any impact to materialize. The contemporaneous effect is estimated to be zero, but after a decade a discovery leads to higher *per capita* GDP growth and an increase in the growth of the urbanization rate.

Anticipation effects Anticipation effects have been shown to be a powerful channel for how discoveries impact the economy (Arezki et al. 2017). While we work with a long time-series, the number of time-series observations is generally limited to one per decade since most of the data we use come from population and economic censuses. In our sample there are only nine instances in which production started in a different decade from the discovery. Looking at these nine cases in a set of auxiliary regressions does not reveal any convincing evidence for an anticipation effect but this might simply be due to a lack of power. As we cannot disentangle anticipation effects due to data limitations, our estimates can be considered as the combined impact of anticipation effects (of both successful and unsuccessful wells) and the production effect.

---

[25] While assigning onshore discoveries to municipalities is straightforward, the mapping is not as clear for offshore discoveries (see Sect. 2). 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.

Human capital We also test for the impact of oil on education, as proxied by (i) mean years of schooling of the population aged 25 years and older, (ii) the share of adults with less than four years of education and (iii) the share of adults with more than 11 years of schooling. The data on education is available from 1970 onwards. We carry out the education analysis using the baseline diff-in-diff specification (from 1970 to 2000).[26] The results are not conclusive: only in one regression do we find a positive impact in the form of a reduction in the share of adults with less than four years of education (see Table A.6 in online Appendix A). Further analysis would be informative but one possible hypothesis is that the increase in growth, urbanization and the services sector, and the associated increase in municipal revenues such as property taxes might have allowed for an increased provision of basic education. Alternatively, urbanization might have directly allowed more children to be able to attend schools, either due to a greater density of schools in urban areas or due to a lower prevalence of child labor (Ersado 2005). Ex-ante one might have expected a negative impact of oil due to a mechanism by which local elites restrict human capital to limit the mobility of workers and to maximize rent extraction (Galor et al. 2009). This mechanism might have offset some of the positive channels so that on aggregate there are no clear results.

Institutions During a large part of the time period we study, Brazil was ruled by highly centralizing dictatorships, allowing for little variation of policies at lower levels of government. After the return of democracy, the 1988 Constitution gave autonomy for municipalities to set taxes and choose expenditures, but there was virtually no variation in institutions among municipalities. To formally test whether institutions differ between oil and non-oil municipalities we use data on institutional variation constructed by Ministry of Planning (2017). Table A.7 in online Appendix A shows no statistically significant difference in institutions between the treated and control groups.[27] We also tested whether discoveries during periods of autocracy had a different effect than discoveries during democracy (see Table A.8 in online Appendix A).[28] We could not find any statistically significant impact.

To study the impact of discoveries on non-oil manufacturing and to understand the impact on the services and agricultural sectors at the worker and firm level in more detail, we turn to matched employer-employee data in Sect. 5.4. We employ a cross-sectional specification, given that the data are not available in the long time series which we used so far. Prior to exploring the underlying mechanisms in that way, however, we check the robustness of our baseline results (Sect. 5.2) and then obtain coefficients which can be interpreted as Local Average Treatment Effect (Sect. 5.3).

5.2 Robustness

Robustness to different specifications In the interest of space we only report tables of the robustness exercises using the baseline dependent variables, but all further results are also robust to the following exercises. The results are both qualitatively and quantitatively robust to using alternative control groups (see Table 9). 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,

---

[26] We also use the cross-section regression specification (only for the year 2000) as in Eq. 4 explained below in Sect. 5.4.

[27] Given that the data do not have a time-series dimension, we use the same cross-sectional specification as in Eq. 4 explained below in Sect. 5.4.

[28] We interact the discovery dummy with a dummy for democracy which is set to 1 for years 1946–1963 and from 1986 onwards.

Table 9 Intention-to-treat effect of oil discoveries: robustness to alternative control groups

Variables	Non-oil municipalities in oil states				Dry drilling, no neighbors			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln GDP per capita	ln population density	ln workers density	Urbanization rate	ln GDP per capita	ln population density	ln workers density	Urbanization rate
Discovery dummy	0.262***	− 0.0560	− 0.114*	0.0519***	0.195**	− 0.0302	− 0.0731	0.0362*
	(0.0781)	(0.0610)	(0.0662)	(0.0190)	(0.0906)	(0.0751)	(0.0822)	(0.0214)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4649	6200	4646	6200	1008	1344	1007	1344
Number of MCAs	775	775	775	775	168	168	168	168
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	FE	FE	FE	FE	FE	FE	FE	FE

Table 9 continued

Variables	All neighbors				Matched neighbors			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	ln GDP per capita	ln population density	ln workers density	Urbanization rate	ln GDP per capita	ln population density	ln workers density	Urbanization rate
Discovery dummy	0.247***	0.0114	−0.0216	0.0434**	0.277***	0.0341	0.0143	0.0419**
	(0.0819)	(0.0641)	(0.0744)	(0.0195)	(0.0863)	(0.0645)	(0.0741)	(0.0206)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1320	1760	1317	1760	768	1024	767	1024
Number of MCAs	220	220	220	220	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	FE	FE	FE	FE	FE	FE	FE	FE

Standard errors clustered at the MCA level. Geographical controls and initial conditions have time-varying coefficients. Discovery is defined as "True Discovery". The initial conditions with time-varying coefficients are: GDP *per capita* in 1949, Urbanization rate in 1940, Worker Density in 1940, and Population Density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

and a matched subsample of adjacent MCAs (matched neighbors). The idea is to create multiple comparison groups to strengthen the results. Overall, the results are remarkably similar across control groups. The estimate for *per capita* GDP ranges from 21.5–31.9% while urbanization is estimated to increase 3.6–5.2% as a consequence of oil discoveries.

Our baseline results are also robust to including log worker density in 1940 as well as 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 (see Table A.9 Columns (1)–(4) in online Appendix A). Moreover, we verify that changing the time period to 1940–1996 does not change the results (see Table A.9 Columns (5)–(8) in online Appendix A). This is important, because it supports the claim that our findings are (mainly) driven by the direct effect of oil production rather than the indirect effect through royalties.

We also tested whether the results are robust to alternative clustering of standard errors and different deflators. Significance of results is highly robust to alternative clustering such as two-way clustering (year and MCA) and spatial clustering (see Table A.10 in online Appendix A).[29] Results are also robust to using alternative prices deflators. In the baseline we use the official GDP deflator constructed by IPEA. As an alternative we deflate nominal GDP using national consumer and producer price indexes.[30]

Robustness to excluding oil and gas processing production facilities. For a sample of U.S. counties, Greenstone et al. (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. We test whether downstream production facilities are driving most of our observed results (as some places with upstream production also have downstream facilities). To evaluate the pure impact of the upstream sector, we exclude those municipalities which host a downstream production facility from both the treatment and the control group and then re-estimate our baseline specification.[31] As can be seen by comparing the last columns in Table A.9 in online Appendix A with the baseline results (from Tables 6 and 7), 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.

5.3 Local average treatment effect

We now turn to estimating the impact of oil production via a 2SLS approach. There are 46 MCAs which have at least one oil production well. As noted earlier, production might be endogenous. We thus estimate a specification similar to Eq. 1 and we instrument for the production indicator (T_{it}) using our discoveries indicator (Z_{it}) to recover a local average treatment effect. Table 10 qualitatively confirms our ITT results. As expected, the estimated coefficients are larger. GDP *per capita* increases by 50%, urbanization by over 6% points, and the share of services workers by over 5% points. Similarly, the impact on sectoral GDP is larger. It is intuitive that the ITT results are scaled up by the proportion of compliers.[32]

---

[29] Conley (1999)'s spatial standard errors are implemented using code provided by Hsiang (2010) and Fetzer (2014). Correlation is assumed to be zero after 1,000 km.

[30] GDP results might be vulnerable to both time-series and cross-section problems with deflators. Since we do not have municipal-level deflators we cannot address the cross-sectional concerns directly.

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

[32] We investigate treatment intensity in Appendix C.

Table 10 Local average treatment effect of oil production

Variables	(1)	(2)	(3)	(4)	Control group: matched dry drilling					
					(5)	(6)	(7)	(8)	(9)	(10)
	ln GDP per capita	ln Population density	ln Workers density	Urbanization rate	ln GDP per capita			Employment share		
					Manufacturing	Services	Agriculture	Agriculture	Public sector	Services
Production dummy	0.411***	−0.0190	−0.0572	0.0644**	0.725**	0.343**	0.105	−0.0608*	0.0137***	0.0522***
	(0.143)	(0.106)	(0.114)	(0.0314)	(0.295)	(0.166)	(0.166)	(0.0320)	(0.00426)	(0.0171)
MCA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	768	1024	767	1024	765	764	765	767	767	767
Number of MCAs	128	128	128	128	128	128	128	128	128	128
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistics	107.24	136.90	153.73	136.90	106.75	106.95	111.30	153.73	153.73	153.73
Estimation	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE

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". The initial conditions with time-varying coefficients are: GDP *per capita* in 1949, Urbanization rate in 1940, Worker Density in 1940, and Population Density in 1940. The geographical controls with time-varying coefficients are: Latitude and Longitude coordinates, dummy for Amazon, and dummy for coastal location.

$*** p < 0.01$, $** p < 0.05$, $* p < 0.1$

5.4 Exploring the mechanism

In this part we investigate the mechanisms underlying our results in more detail. We aim to shed light on three questions related to the structural transformation occurring in the local economy due to oil discoveries: (i) What exactly happens to the services sector? (ii) What happens to the 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 from Ministry of Labor's RAIS (Relação Anual de Informações Sociais). The RAIS dataset has information on each formal worker at each plant in Brazil. One key aspect which we will exploit here is that RAIS looks only at formal workers, while the population and employment census data which we exploited in the panel analysis above include both formal and informal workers. We complement the RAIS data with cross-sectional data on informality from the 2000 population census microdata, collected by the Brazilian Bureau of Statistics. We use 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.

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 Sect. 4 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:

$$Y_i = \alpha + \tau_{cs} Z_i + \beta' X_i + \epsilon_i, \tag{4}$$

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.

The first four columns of Table 11 show that the cross-sectional results are in line with previous findings: In 2000, the assigned to treatment group has a higher *per capita* GDP and is more urbanized, but population density and total worker density are not affected by oil discoveries. Places where oil was discovered do have a higher *formal* worker density and a higher share of workers in the formal sector (Columns (5)–(6)).[33] Furthermore, average wages are higher in oil discovery municipalities while firm density is not statistically different between the discovery and control groups.

Columns (9)–(14) investigate which formal sectors are affected by oil discoveries. 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. We find that the formal manufacturing sector (excluding natural resource extraction) and the formal 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. The formal agricultural sector also does not seem to be affected. By contrast, the growth in the number of formal workers is driven by an increase in the number of formal workers in services.

---

[33] In our sample, the informal sector is large: On average, only 35% of workers are formally employed. A large fraction of informal labor occurs in the agricultural sector. The definition of formal employment is from the Brazilian Bureau of Statistics and includes workers with a valid work card, those who work in the military or judiciary, and self-employed workers who contribute to social security.

Table 11 Mechanisms: oil discoveries, wages, worker density, and firm density in 2000

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Baseline results				Control group: matched dry drilling				Firm density (formal sector)		
	ln GDP per capita	ln population density	ln worker density	Urbanization rate	ln Worker density formal sector	% Workers in the formal sector	ln Firm density	ln Average wage	Manufacturing	Services	Agriculture
Discovery dummy	0.396***	−0.0269	−0.0621	0.0551*	0.506*	4.352**	0.384	0.185**	0.308	0.426	0.274
	(0.120)	(0.129)	(0.136)	(0.0301)	(0.287)	(1.710)	(0.285)	(0.0739)	(0.338)	(0.302)	(0.286)
Observations	128	128	128	128	128	128	128	128	121	128	122
Number of MCAs	128	128	128	128	128	128	128	128	121	128	122
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Table 11 continued

Variables	Control group: matched dry drilling									
	Worker density (formal sector)			Service sector						Agriculture
	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
	Manufacturing	Services	Agriculture	ln firm size	ln skilled worker density	ln unskilled worker density	Skilled worker fraction	ln avg. skilled wage	ln avg. unskilled wage	Fraction total workers
Discovery dummy	0.338	0.796**	0.546	0.370***	0.711*	0.685**	−0.0188	0.168***	0.0860	−0.056**
	(0.450)	(0.353)	(0.359)	(0.118)	(0.363)	(0.346)	(0.0260)	(0.0793)	(0.0611)	0.027
Observations	121	128	122	128	128	127	128	128	127	128
Number of MCAs	121	128	122	128	128	127	128	128	127	128
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

This table shows the results of estimating Eq. 4. 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. The pre-treatment variables are urbanization rate in 1940, population density in 1940, worker density in 1940 and *per capita* GDP in 1949. The geographical controls are Latitude and Longitude coordinates and indicator variables showing whether the MCA is located in the semiarid region, in the Amazon region, or on the coast.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Columns (15)–(20) further disaggregate the data for services. 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 number 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.[34] Lastly, Column (21) shows that the fraction of total (formal and informal) workers employed in agriculture is lower in oil municipalities.

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. Given the large pool of workers in the informal agricultural sector (in 1940 roughly 70% of workers worked in agriculture on average in the sample of municipalities which drilled for oil), the elasticity for unskilled workers appears to be so high that more workers can be attracted at virtually no higher pay. Looking at the distribution of wages, we find that the average low skilled formal services sector wage in 2000 in our sample was above the value of the national monthly minimum wage. Given the very low income an informal subsistence farmer could expect, it seems that the opportunity to work at or above the legal minimum wage was a sufficient incentive to move to the urban agglomeration and enter the formal workforce.

More broadly, it is useful to quantify the results. The increase in the share of services workers we estimated translates to roughly half a standard deviation of the distribution in the sample in 2000. Similarly, the estimated 45% increase in the size of services firms corresponds to a quarter of the standard deviation of the distribution in 2000. The mean size was 9 workers, implying on average 4 additional workers per firm due to oil discoveries. Skilled wages are estimated to be about close to 20% higher in discovery MCAs. This corresponds to 150 Real per month — equivalent to the value of the national monthly minimum wage in 2000.[35]

5.5 Discussion

Our core result is that the positive demand shock associated with oil discoveries and ultimately oil production leads to a reallocation of labor from (subsistence) agriculture to urban services. During our period of analysis, Brazil still had a large subsistence agricultural sector, which employed a substantial number of people with very low productivity. This shift of employment led to an increase in urbanization and higher GDP *per capita* due to higher marginal productivity in the services sector. Since at the same time we do not find any significant increase in population density or any spillovers to neighboring municipalities, the result implies an important spatial segmentation of labor markets. Furthermore, since we do not find a negative impact on output in the agricultural sector, it must be that (i) marginal productivity of labor in the agricultural sector was very low indeed so that the impact of the reallocation on agricultural output is unimportant/undetectable and/or (ii) local price effects not captured by the national deflators offset a potential reduction in agricultural output.

---

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

[35] Table A.11 in Appendix A shows the microdata regressions employing using a 2SLS estimator. As before, in the 2SLS approach, oil production is instrumented using oil discovery. All results are qualitatively the same but estimated coefficients are larger.
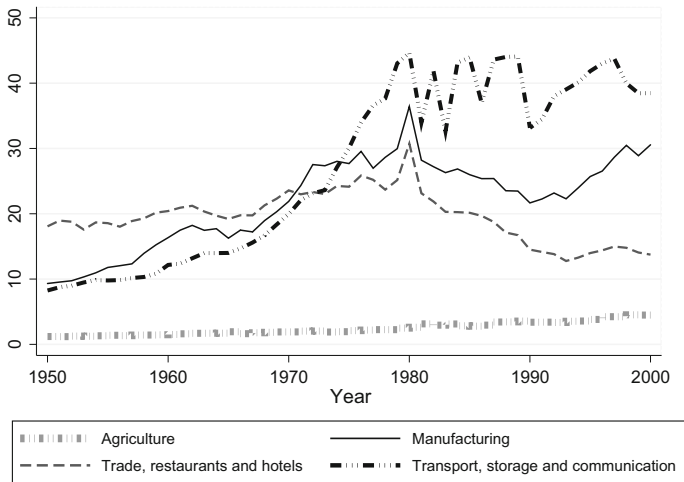
Fig. 5 Sector labor productivity: 1950–2000. *Notes* Figure shows the labor productivity for four sectors in Brazil during the period from 1950 to 2000. Data are from Timmer et al. (2015). The sectors are: (i) agriculture, (ii) manufacturing, (iii) trade, restaurants and hotels and (iv) transport, storage and communication. Labor productivity is calculated by dividing the value added at constant 2005 national prices (in millions) and persons engaged (in thousands)

Several studies on local shocks and their impact on labor mobility support the notion of limited interregional labor mobility. For Brazil, Dix-Carneiro and Kovak (2017b) show evidence of imperfect interregional labor mobility after a negative labor demand shock (brought about by a trade policy reform).[36] Dix-Carneiro and Kovak (2017a) also find minimal effects of regional shocks on inter-regional worker mobility; they find that the informal sector acts as the margin of adjustment. In our paper, the subsistence agriculture sector provides the margin of adjustment. Low mobility across regions after commodity shocks is also found in other countries in Latin America, e.g., Chile (Pellandra 2015). Manning and Petrongolo (2017) present a model of job search behavior across space and show that local shocks in isolated areas have larger effects as local stimuli are not dissipated to other areas. Last, it is also worth pointing out that our unit of analysis (MCAs) is not small. Since 1940 MCAs comprise several municipalities, they can be broadly thought of as an approximation of local labor markets.[37]

In terms of productivity in the agricultural sector, several elements suggest that the regions in Brazil studied in our paper were very close to a Lewis-type world in which agricultural marginal labor productivity was close to zero. Figure 5 uses data from Timmer et al. (2015) to plot sectoral labor productivity between 1950 and 2000 for Brazil. A large gap between productivity in agriculture and other sectors persisted during the past several decades. Especially in the middle of the 20th century, labor productivity in agriculture was very low and stagnant. Only more recently technological developments in Brazilian agriculture (e.g., the introduction of genetically engineered soybean seeds) have improved the productivity in this sector (Bustos et al. 2016).

---

[36] Dix-Carneiro (2014) estimates high interregional mobility costs in Brazil, with the median value ranging from 1.4 to 2.7 times annual average wages.

[37] Due to data constraints (our data come from decennial censuses) we cannot observe seasonal workers and we thus cannot rule out temporary migration. Given that we do not find any spillovers to municipalities which neighbor discovery municipalities, however, these seasonal workers would have to be based far away from their place of work, perhaps reducing the likelihood that this takes place in larger numbers.

Our results are consistent with the framework presented by Gollin et al. (2016), who show the possibility of natural resource-led growth of "consumption cities". In their model of labor allocation between rural and urban areas, rural production consists only of food, which is close to our setting in Brazil. Urban production consists of non-food tradables and non-tradables. In their model, the income effect of an expansion of the resource sector leads to urbanization as relative demand shifts towards urban goods. A standard Dutch Disease effect leads to further reallocation from both the food and urban tradable sectors to the urban non-tradable sectors. Nearly all these predictions are consistent with our results of reallocation from subsistence agriculture to urban services leading to urbanization. While Gollin et al. (2016) assume equal productivity levels across sectors, assuming instead a higher productivity in urban services than rural food (as in Fig. 5) would yield higher GDP in oil municipalities.

Traditional theoretical frameworks which incorporate a natural resource sector are less suited to study the context of Brazilian municipalities. This is largely because of the absence of a rural/food sector. In Corden and Neary (1982), for example, an increase in resource wealth yields an expansion of the non-tradable sector (as in our results), a real appreciation of the exchange rate and, in contrast to our results, a crowding out of the non-resource tradable sector. If the tradable sector benefits from learning-by-doing but the non-tradable sector does not (or less), then aggregate growth is permanently reduced by the natural resource expansion (Krugman 1987; Matsuyama 1992). Since in our results the non-tradable sector draws labor from the (low productivity) rural food sector, a framework which instead focuses on the crowding out of a (high productivity) sector is likely to predict aggregate results which are more negative than those we find.

Notice that even in Gollin et al. (2016) there is some reallocation from the urban tradable to the urban non-tradable sector. We can only hypothesize why we do not observe this in our data. One possibility is that the small initial size of the manufacturing sector makes it hard to detect any negative impact. We return to this in the below discussion of external validity.

External validity Broadly speaking, local level results of oil discoveries should be similar if the economic environment that we work with—regarding labor mobility and the presence of a low productivity sector with ample labor supply—are satisfied in other countries. A comparison with results obtained by Allcott and Keniston (2018) for the U.S. highlights the importance of initial economic conditions for the impact of oil. Allcott and Keniston (2018) present a model where the interplay of agglomerative effects and the crowding out effect of a resource boom can lead to either an increase or reduction in welfare. Looking at local economies in the U.S., they provide evidence that oil booms benefit oil-linked subsectors of manufacturing but lead to a contraction of the tradable manufacturing subsector (total factor productivity of the latter is unaffected, however).

The fact that we do not find any impact on the non-oil manufacturing sector—neither positive nor negative—is thus likely to be specific to the particular situation of a developing country with relatively little large-scale manufacturing in the affected regions. By contrast, in the U.S. positive spillovers from oil production in local economies were possible given that an important nucleus of manufacturing, with strong input–output linkages to the oil sector, existed.[38]

Apart from the economic environment, our identification strategy and the institutional environment could potentially also restrict the external validity of the results. In terms of the identification strategy, our results are derived conditional on drilling taking place. Given that

---

[38] We interact the treatment dummy with the baseline manufacturing or agriculture intensity of the local economy to obtain a specific test of potentially heterogenous treatment effects based on different initial conditions. The regressions are under-powered to detect such heterogenous effect, however. The results are available upon request.

oil drilling is not random, this raises the question of whether the results apply more broadly. However, as shown in Table A.12 in online Appendix A, looking at all MCAs there seems to be no correlation between economic characteristics at the beginning of the sample period and subsequent drilling attempts — there is only a correlation with geographical characteristics. And while our preferred identification strategy derives results conditional on drilling, we have shown (Table 9) that the results are broadly unchanged when we use larger sets of Brazilian MCAs as the control group.

The geographical distribution of oil discoveries and production in Brazil is broadly similar to that in other large oil-producing countries. In Brazil, the U.S. and Russia, (i) oil does not come only from one part of the country but (ii) total production volumes are highly concentrated in a few regions. In our sample, production takes place in 46 MCAs (distributed among 9 out of 26 states in Brazil), while the top 10 producing MCAs account for 90% of production in 2000. In the U.S., counties in different states hold oil and gas reserves, but crude oil production in Texas accounted for more than 35% of total U.S. production in 2016. Similarly, in Russia different regions produce hydrocarbons, but the top producers account for a large fraction of total production. The fact that a few mega-fields account for a large share of national production does not rule out that levels of production that are relatively small for national economies matter for local economies.

As for the institutional environment, our estimates come from a setting where subsoil assets are owned by the state and a national oil company drills for oil across the country by identifying municipalities which are geologically-suitable for oil drilling and discoveries. This is in fact not an uncommon setup around the World. In most countries subsoil assets are property of the state and it is estimated that national oil companies control 90% of world oil reserves and over 70% of production (Venables 2016). The U.S. is different as it has more widespread ownership of resources than Brazil. Besides, there are thousands of oil companies in the U.S., in contrast to the historical monopoly of Petrobras in Brazil. Nevertheless, the mechanisms studied in this paper do not rest on the particular institutional framework. As long as the conditions on the economic structure explained above hold and rents accrue outside of the local economy, the same mechanisms would be expected to apply, regardless of whether a state monopoly or a private player is involved.

6 Conclusion

We investigated the effects of natural resource extraction on local economic development in a developing country using a quasi-experimental identification strategy and documented a positive growth effect of oil discoveries. We found evidence of a structural transformation towards the urban services sector, as identified by a positive impact of oil discoveries on urbanization as well as increases in services GDP, share of workers in services, and the size of services firms.

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, nominal appreciation and pork barrel politics). We show that at the local level, oil discoveries are not a curse per se, and in the context of a developing country, the pure market effect (i.e., when fiscal windfalls are small) 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 for developing countries.

References

Acemoglu, D., Finkelstein, A., & Notowidigdo, M. J. (2013). Income and health spending: Evidence from oil price shocks. *Review of Economics and Statistics*, *95*, 1079–1095.

Allcott, H., & Keniston, D. (2018). Dutch disease or agglomeration? The local economic effects of natural resource booms in Modern America. *The Review of Economic Studies*, *85*, 695–731.

ANP. (2012). Anuário Estatístico Brasileiro do Petróleo, Gás natural e Biocombustíveis, Anuário estatístico da ANP, Agência Nacional do Petróleo, Gás Natural e Biocombustíveis.

Aragón, F. M., & Rud, J. P. (2013). Natural resources and local communities: Evidence from a Peruvian gold mine. *American Economic Journal: Economic Policy*, *5*, 1–25.

Arezki, R., Ramey, V. A., & Sheng, L. (2017). News shocks in open economies: Evidence from giant oil discoveries. *The Quarterly Journal of Economics*, *132*, 103–155.

Brollo, F., Nannicini, T., Perotti, R., & Tabellini, G. (2013). The political resource curse. *American Economic Review*, *103*, 1759–96.

Bustos, P., Caprettini, B., & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, *106*, 1320–65.

Caselli, F., & Michaels, G. (2013). Do oil windfalls improve living standards? Evidence from Brazil. *American Economic Journal: Applied Economics*, *5*, 208–38.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, *92*, 1–45.

Corden, W. M., & Neary, J. P. (1982). Booming sector and de-industrialisation in a small open economy. *The Economic Journal*, *92*, 825–848.

Cotet, A. M., & Tsui, K. K. (2013). Oil and conflict: What does the cross country evidence really show? *American Economic Journal: Macroeconomics*, *5*, 49–80.

Cotet, A. M., & Tsui, K. K. (2013). Oil, growth, and health: What does the cross-country evidence really show? *The Scandinavian Journal of Economics*, *115*, 1107–1137.

Cust, J., & Harding, T. (2014). *Institutions and the location of oil exploration*. Technical report 127, OxCarre Research Paper.

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

Davis, D. R., & Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, *92*, 1269–1289.

Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica*, *82*, 825–885.

Dix-Carneiro, R., & Kovak, B. K. (2017). *Margins of labor market adjustment to trade*. Durham: Mimeo, Duke University.

Dix-Carneiro, R., & Kovak, B. K. (2017). Trade liberalization and regional dynamics. *American Economic Review*, *107*, 2908–46.

Dube, O., & Vargas, J. (2013). Commodity price shocks and civil conflict: Evidence from Colombia. *The Review of Economic Studies*, *80*, 1384–1421.

Ersado, L. (2005). Child labor and schooling decisions in urban and rural areas: Comparative evidence from Nepal, Peru, and Zimbabwe. *World Development*, *33*, 455–480.

Fetzer, T. (2014). *Can workfare programs moderate violence?* Evidence from India, economic organisation and public policy discussion papers, eopp 53. The London School of Economics and Political Science, Suntory and Toyota International Centres for Economics and Related Disciplines.

Fustier, K., Gray, G., Gundersen, C., & Hilbott, T. (2016). *Global oil supply: Will mature field declines drive the next supply crunch?*. Report, HSBC Global Research.

Galor, O., Moav, O., & Vollrath, D. (2009). Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. *The Review of Economic Studies*, *76*, 143–179.

Gollin, D., Jedwab, R., & Vollrath, D. (2016). Urbanization with and without Industrialization. *Journal of Economic Growth*, *21*, 35–70.

Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, *118*, 536–598.

Harbaugh, J., Davis, J., & Wendebourg, J. (1995). *Computing risk for oil prospects: Principles and programs.*, Computer methods in the geosciences Amsterdam: Elsevier Science.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences*, *107*, 15367–15372.

Idson, T. L., & Oi, W. Y. (1999). Workers are more productive in large firms. *American Economic Review*, *89*, 104–108.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*, 5–86.

Kellogg, R. (2011). Learning by drilling: Interfirm learning and relationship persistence in the Texas oilpatch. *The Quarterly Journal of Economics*, *126*, 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*, 41–55.

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

Manning, A., & Petrongolo, B. (2017). How local are labor markets? Evidence from a spatial job search model. *American Economic Review*, *107*, 2877–2907.

Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth. *Journal of Economic Theory*, *58*, 317–334.

Michaels, G. (2011). The long term consequences of resource-based specialisation. *Economic Journal*, *121*, 31–57.

Ministry of Planning. (2017). *Agenda Politico-Institucional*. Technical report, Brazilian Ministry of Planning

Monteiro, J., & Ferraz, C. (2012). Does oil make leaders unaccountable? Evidence from Brazil's offshore oil boom, mimeograph, PUC-Rio.

Pellandra, A. (2015). The commodity price boom and regional workers in Chile: A natural resources blessing? Mimeo, Economic Commission for Latin America and the Caribbean.

Reis, E., Pimentel, M., & Alvarenga, A. (2007). Áreas Mínimas Comparáveis para os Períodos Intercensitários de 1872 a 2000, mimeograph, Ipea.

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

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169–188.

Sachs, J., & Warner, A. (2001). The curse of natural resources. *European Economic Review*, *45*, 827–838.

Stefanski, R. (2014). Structural transformation and the oil price. *Review of Economic Dynamics*, *17*, 484–504.

Timmer, M. P., de Vries, G. J., & de Vries, K. (2015). Patterns of structural change in developing countries. In J. Weiss & M. Tribe (Eds.), *Routledge handbook of industry and development* (pp. 65–83). Abingdon: Routledge.

van der Ploeg, F. (2011). Natural resources: Curse or blessing? *Journal of Economic Literature*, *49*, 366–420.

Venables, A. J. (2016). Using natural resources for development: Why has it proven so difficult? *Journal of Economic Perspectives*, *30*, 161–84.