# Oil discoveries and economic development in the US, 1880 to 1910

Nils Braakmann

*Newcastle University Business School, NE1 7RU, Newcastle upon Tyne, UK.*

No 2011/07

Newcastle Discussion Papers in
Economics: ISSN 1361 - 1837

# Oil discoveries and economic development in the US, 1880 to 1910

Nils Braakmann*

Newcastle University

[This version: September 8, 2011]

**Abstract:**

This paper examines the economic consequences of the US oil boom from 1880 to 1910. I combine information on oil discoveries with US census data and investigate the effects on various indicators of development. Evidence from OLS and IV-estimates indicate that oil discoveries (a) increased urbanization, (b) drew more (male) workers into mining and manufacturing, away from agriculture, (c) improved the economic situation of men and women and (d) improved the education of men and women. Effects are generally weaker for women, which supports the idea that changes in employment opportunities were the main driver of these changes.

**Keywords:** resource curse, Dutch disease, oil boom

**JEL Classification**: J21, J24, N31, N51, N61, O13, O14, Q32, Q33, R11

**Word count:** 7986

---

* Nils Braakmann, Newcastle University, Business School – Economics, Newcastle upon Tyne, Tyne and Wear, NE1 7RU, United Kingdom, e-mail: nils.braakmann@ncl.ac.uk.

All analyses used Stata 11.1. Do-Files are available from the author on request. All analyses and opinions expressed in this paper as well as any possible errors are under the sole responsibility of the author. I thank Laura Hanson for help with the GIS data.

## I. Introduction

Oil has arguably played a large role in American history. This paper presents the first microeconometric study on the effects of oil discoveries during the early period of the American oil boom. I focus on the period from 1880 to 1910, i.e., between the end of the American Civil War, the subsequent Reconstruction of the South and the beginning of World War I. During this period major oilfields were discovered in Indiana, Ohio, Kansas, Oklahoma, Arkansas, North Louisiana and parts of Texas. Furthermore, the commercial value of oil was already well recognized at this time, which also saw the heyday of Rockefeller's *Standard Oil*.

I combine information on the location and time of discovery of all known oilfields in the US with IPUMS micro data from the 1880 and 1910 US censuses[1] to look at the effects the discovery of oil had on urbanization, education, sectoral employment patterns as well as on various labor market outcomes. The underlying logic is that the discovery of oil changed incentives for work and education, drawing individuals into the oil industry and into acquiring the skills necessary for working in that industry. These changes in the allocation of labor and the changed incentives for human capital acquisition can influence long-term growth prospects leading to the phenomenon

---

[1] Due to a fire in 1921, which destroyed the 1890 census records, no micro data is available for 1890. An earlier version of this paper additionally used the 1900 census and looked at developments from 1880 to 1900 and from 1900 to 1910. As results very qualitatively identical to those from 1880 to 1910 and very similar in both periods, the present paper focuses on 1880 to 1910.

known as “resource curse” (Sachs and Warner, 1995, 2001; see van der Ploeg, 2011, for an overview) or “Dutch disease”.

Classical explanations for these effects include economic as well as political channels. Among the first group are changes in factor prices caused by the development of a resource intensive sector, which may influence (and typically hamper) developments in other sectors of the economy (see, e.g., Corden and Neary, 1982), changes in the incentives to invest in education – typically thought of as negative incentives as most resource intensive sectors are relatively low-skilled compared with, e.g., manufacturing – (see, e.g., Leamer et al., 1999; Gylfason, 2001) and finally increased inequality when the returns of the natural resource benefit only a relatively small group of individuals (e.g., Leamer et al., 1991; Humphreys et al., 2007). Political explanations (see Ross, 1999, for an overview) include increased corruption, an increase in rent-seeking behavior (see, e.g., Tornell and Lande, 1999; Mehlum et al., 2006a,b), overconfidence by political decision-makers leading to a neglect of growth-enhancing policies as well as increases in political instability leading as far as civil wars (see, e.g., Caselli and Coleman II, 2006; Ross, 2006). In a recent paper James and James (2011) explore the alternative explanation that the slower growth of resource abundant regions can be explained by lower growth rates in the mining sector, while resource extraction still leads to real income increases.

It is important to stress that many of these explanations have been written with current growth differences between resource abundant and resource poor countries like the differences between the oil-rich states in the Middle East and the Western industrialized countries in mind. This fact calls for some caution to directly apply the same reasoning to historical developments like the 19th century oil boom in the US. It is, for instance, not entirely clear whether there was indeed a negative incentive to invest in education in the US, as oil production required higher skills than previously dominant agriculture.

In the following I test some of these commonly made predictions. I will not deal with certain explanations for the resource curse that operate on the level of the whole economy, e.g., through exchange rate changes, as these are impossible to test with data from just one country. I will also not deal with the political explanations for the resource curse due to a lack of suitable data. I will, however, investigate changes in urbanization, shifts in sectoral employment, changes in labor market participation rates and the economic situation of working individuals.

To the best of my knowledge this paper is the first microeconometric study of the American oil boom. There is, however, a closely related paper by Michaels (2011) that focuses on the long-term growth effects of oil abundance in the American South. His results, relating to the period 1940 to 1990, show that oil abundant counties fared better than non-oil abundant counties in terms of a number of economic characteristics throughout most of the 20th century although most differences had disappeared by 1990.[2] An earlier paper by Papyrakis and Gerlagh (2007) found

[2] There are also some other papers that investigate the occurrence of a resource curse using differences in resource abundance within countries. Examples include Caselli and Michaels (2009) and Naritomi et al. (2007) for Brazil and Vicente (2010) for a comparison of Sao Tome and Principe and Cape Verde after oil was found in the former. Aragón and Rud (2011) find positive effects of a single Peruvian gold mine

evidence that resource abundance slowed growth in US states between grew slower from 1986 to 2011, a fact they explain by decreased investment, lower levels of schooling, openness and R&D expenditure and increased corruption.

This paper differs from these earlier studies in a number of important aspects. Firstly, while Michaels (2011) is effectively looking at the availability of oil and its long-term growth effects and while Papyrakis and Gerlagh (2007) look at recent years, I focus on the historical events surrounding the discovery of oil. In other words, while the other studies look at the effects of oil abundance on economic outcomes from 1940 onwards, my paper focuses on the (short to medium term) effects of the changes in incentives and economic outcomes brought forth by the abrupt discovery of a natural resource. Secondly, the fact that I use micro data from the 1880 and 1910 US censuses allows me to study a different set of outcome variables, including urbanization, school attendance, literacy as well as various labor market outcomes. It also enables me to look at differential effects by gender, which turn out to be important. Thirdly, I apply two identification strategies to tease out the causal effects of the oil and gas discoveries: The first, which is similar to the approach used by Michaels (2011), is a difference-in-differences-type strategy. Here, I treat the discovery of oil in a county as an exogenous shock to individuals in that county and compare the change in treated counties with changes in untreated counties while controlling for a number of factors including state-year-trends. This strategy arguably

---

on households in the surrounding area. The classical studies of the resource curse generally relied on cross-country comparisons, e.g., Sachs and Warner (1995), Leite and Weidmann (1999), Isham et al. (2005), Kolstad (2009), Collier and Goderis (2009) and Brunnschweiler and Bulte (2009).

does not account for the fact that the timing of oil discoveries might still be driven by unobserved factors on the county level that could also influence the outcomes. To overcome this potential endogeneity problem, I rely on an instrumental variable strategy. Specifically, I use the fact that the discovery of oil spread out towards the South West and the North East from Oil Creek in Venango County, Pennsylvania, where the first commercial oil well was drilled in 1859. This historical pattern allows me to use the distance between a county and this first oil well as an instrument for the discovery of oil in a given period. This instrument is similar in spirit to Becker and Woessmann (2009) who instrumented the spread of Protestantism through Prussian counties by the county's distance from Wittenberg. In both cases a historical event happened (coincidently) in some place (Luther's theses in Wittenberg, the first successful drilling in Venango County) and spread out from there. As I show in section 3, counties within the same state that are closer or farther away from Venango County generally do not differ in their characteristics in 1880, which point towards the validity of the IV strategy/

My results indicate that that oil discoveries (a) increased urbanization at the expense of individuals living on farms, (b) led to sectoral shifts by drawing more (primarily male) workers into mining and manufacturing and away from agriculture, (c) improved the economic situation of men and women by drawing them into occupations that enjoyed a higher socio-economic status, and (d) broadly improved the education of men and women, which is likely related to the shifts in employment. The effects are generally somewhat weaker for women, which supports the idea that changes in employment opportunities were the main driver of these changes.

Section II presents some historical background information, Section III introduce the data and the estimation strategy, results can be found in Section IV, Section V concludes.

## II. Historical background

This section briefly sketches the history of the US petroleum industry as far as it is relevant for this paper. A full historical description can be found in Owen (1975), a contemporaneous view on the same subject is Bacon and Hamor (1916).

The commercial extraction of petroleum in the United States is generally considered to have begun in 1859 with the Drake Well in Oil Creek, Vernango County, Pennsylvania. Oil was already extracted prior to that date, either as coal oil or, in the case of petroleum, as a by-product of wells originally drilled for the production of salt (see, e.g., Bacon and Hamor, 1916, pp. 197-212).

The fact that the first successful drilling occurred in Venango County can be considered a historical coincidence. Edwin Drake, a former train conductor, was hired in the spring of 1858 by the Seneca Oil Company to investigate possible oil deposits in Titusville, PA. The actual drilling process started in 1859 using techniques from the drilling of salt wells, but originally went slowly and encountered technical difficulties due to collapsing boreholes. Seneca Oil largely abandoned the project at this time and drilling only continued due to Drake being helped financially by friends and some local merchants. Fate turned when Drake invented a new drilling technique, using an iron pipe to stabilize the borehole with the actual drill working inside this pipe. On August 27, 1859, Drake’s drill struck oil, which triggered a series of new drillings, leading to the establishment of commercial enterprises and increasing the oil production in Pennsylvania from 2,000 barrels in 1859 to 500,000 barrels in 1860 (Owen, 1975, p. 12). During this early period oil was mainly used in lamps and increasingly in the production of lubricants that were used in various industries (Owen, 1975, p. 12).

Following the initial successful drilling of the Drake Well, exploration activities quickly spread into other parts of the US, leading to the discoveries of large oilfields in the Appalachian Basin. At the end of the American Civil War, oil was extracted in Pennsylvania, New York State, Ohio, Kentucky and today's West Virginia as well as in some fields in the Rocky Mountains. In the early 1880s a large gas field that was later found to also contain a great amount of oil was found in Indiana. Around that time, oilfields were also discovered in California (first discovery of a large field in 1880[3]), Kansas (1892), Texas (1894), Oklahoma (1897) and Louisiana (1906). Figure 1 provides a plot of these discoveries. Note that the empirical analysis will use county-level discoveries, i.e., even states were oil was discovered prior to 1880 will contribute to the empirical results.

(Figure 1 about here.)

The discovery of oil typically had a transforming effect on the respective region, in particular leading to rapid urbanization and migration for the period of extraction and

[3] There was some extraction of oil in California prior to 1880, but the output from these wells was generally negligible. After 1881 California soon became one of the states with the highest petroleum production (Bacon and Hamor, 1916, p. 235).

subsequent decline when the oil reserves were exhausted (see Michaels, 2011, for evidence on long-term migration into oil abundant counties). As an example, Bacon and Hamor (1915, pp. 223-225) report on the discovery of oil at Pithole:

> "Pithole City was a typical "oil country" city. Built up in an incredibly short time, it had a population estimated at 14,000 before the end of September, 1865, and its post office ranked next in importance in the State to those of Philadelphia and Pittsburgh. As, however, its production fell off, its prosperity rapidly declined, and within two years of its foundation it was practically deserted. Throughout the "oil country," as the producing fields changed, the population shifted with the fields, and the towns that had sprung from the wilderness vanished almost as quickly as they had grown."

A further aspect to keep in mind is that many of the counties were oil was discovered were predominantly agriculturally oriented (see Wright, 1986; Michaels, 2011). This has certain consequences for the analysis conducted here as, e.g., the oil industry was more likely to crowd out agriculture than (non-existent) manufacturing. As the oil industry was also relatively skill-intensive compared with agriculture, we might expect to see positive incentives for acquiring education, which runs contrary to common perceptions of the resource curse (see also Michaels, 2011, for evidence from 1940 onwards).

To gain an idea about the economic importance of oil at the end of the $19^{th}$ century, Figure 2 plots the price of oil in 2000 prices over the period 1860 to 2010 (see panel (a)). In the early years of the petroleum industry and during the American Civil War, prices were quite volatile and could reach heights that were not seen again until the 1980s. From 1880 onwards a period of relative price stability began that lasted for almost 100 years until the oil crisis of the early 1970s. The important implication for this paper is that the worth of oil was already well recognized during the observation period. Furthermore, Figure 1 also suggests that changes in oil prices can safely be ignored for the purpose of this paper, as these were essentially constant (see panel (b) for a more detailed view).

(Figures 2 and 3 about here.)

Figure 3 presents an explanation for this relative price stability, which initially seems counterintuitive in the face of rising demand through the beginning proliferation of automobiles and an increased industrialization (see Wright, 1990, for an analysis of American industrialization and the role of natural resources from 1879 onwards), namely a large rise in supply. The domestic crude oil production in the US increased each year until 1970. Increases were stronger after 1900 when many of the larger oilfields were found, the extraction technologies improved and public policies towards reducing wastefulness were brought into place (see, e.g., Davidson, 1963, Libecap and Wiggins, 1984, or Libecap, 1989, for economic analyses of some of these policies).

## III. Data and estimation

In this paper I use two sources of data. Information on oilfields comes from the *Oil and Gas Field Code Master List 2008* (Energy Information Administration, 2009), which is published by the statistical agency within the US Department of Energy. This list contains the location and year of discovery for every oilfield in the United States. An oilfield is defined as "an area consisting of a single reservoir or multiple reservoirs all grouped on, or related to, the same individual geological structural feature and/or stratigraphic condition. There may be two or more reservoirs in a field which are separated vertically by intervening impervious strata, or laterally by local geologic barriers, or by both." (Energy Information Administration, 2009, p. 2) Fields covering multiple counties and/or states are assigned to each of these.

There are a couple of remarks that have to be made on this data. First, unlike Michaels (2011) I focus on oil discoveries instead of oil availability. The reason for this difference in approaches is that I am primarily interested in the change in incentives and subsequent behavior brought forth by the beginning resource extraction. For these changes in incentives to happen, the local population needs to know about the resource in the first place. Second, while the data contains information on offshore and onshore oilfields, the former are irrelevant in the context of this paper as the earliest discovery of an offshore field takes place in 1922, i.e., 12 years after the end of the observation period in this paper.

This information on oil fields is merged with individual level data from the 1880 and 1910 IPUMS public use census data files (Ruggles et al., 2010). Specifically, I use the coordinates of the centroid of the county the oilfield is located in in 2008 and merge each oilfield to the closest county in both 1880 and 1910, where "closest" is defined in terms of the geodetic distance between the counties' centroids.[4][5]

---

[4] All distance calculations in this paper use the ado-files *geonear* and *geodist* by Robert Picard.

[5] Note that using a naïve matching approach based on the 2008 FIPS codes – essentially ignoring all changes to counties during the last century – does not lead to substantially different results.

In this paper, I focus exclusively on whites. Given small sample sizes for most ethnicities, this restriction substantially amounts to omitting blacks. The main reason for this is the very complicated situation of blacks in a period shortly after the abolishment of slavery. Dealing appropriately with the multitude of formal and informal institutions in various states and how these might influence the relationship between resource discoveries, rent sharing and the various outcome variables is well beyond the scope of this paper. Restricting the sample in this way leads to sample sizes of 639,769 men and 606,690 women, of which 85,577 men and 79,890 women live in a country with at least one oilfield in the respective year.

I measure urbanization by two indicators, specifically whether an individual lives in an urban area (defined as cities and incorporated places with 2,500+ inhabitants) and whether the individual lives on a farm. Human capital is measured by a dummy for an individual currently attending school and another dummy for being illiterate. Note that there is no information on degrees or years of education available in the data. Sectoral shifts are measured by dummies for being in a mining related occupation (including miners, mining engineers and geologists), for being an agricultural worker and for being a manufacturing worker. Furthermore, I also consider the question whether more individuals were drawn into the labor force. Note that during the period considered in this paper unemployment insurance/benefits were not known yet (the first state to introduce them was Wisconsin in 1932), hence there is no distinction between employed and unemployed individuals in the data. Finally, as the data does not contain information on income, I proxy changes in the economic situation of individuals the “occupational income score” provided by IPUMS. In essence, this score assigns each occupation the median income of individuals in this occupation in 1950 (measured in hundreds of 1950 Dollar).[6] From the perspective of this paper, this variable provides information on whether the discovery of oil led to more people working in occupations with higher future wages.

(Table 1 about here.)

Table 1 present basic descriptive statistics for the pooled samples and split up by year. Looking at trends over time, we can see (a) an increase in urbanization and a decrease in the share of individuals living on farms, (b) an increase in human capital, (c) a shift away from agriculture towards manufacturing and (on a lower level) mining, (d) an increase in female labor market participation and (e) general improvements in the economic conditions of individuals as measured by the occupational income score. The following analyses will deal with the question whether any of these developments were stronger or weaker in counties where oil was discovered than in other counties.

In a first step, I run regressions of the form

$$y_{icst} = \alpha + \phi_{st} + \eta_c + \beta' X_{icst} + \tau * Oil_{ct} + \varepsilon_{icst}, \qquad (1)$$

where $y_{icst}$ is the respective outcome of individual *i* observed in county *c* (nested within state *s*) in year *t*. Note that the individual level data are repeated cross-sections. $\eta_c$ is a county level fixed effect, $X_{icst}$ contains strictly exogenous individual level control variables, specifically dummies for being a first or second generation migrant and age as a second order polynomial. $\phi_{st}$ are state-year specific fixed effects that

---

6 See http://usa.ipums.org/usa/chapter4/chapter4.shtml#occscore for details on how this variable was constructed.

capture any state wide unobserved changes, e.g., the introduction of compulsory schooling laws during the period under investigation here, which may influence some of the outcomes like school attendance.[7] The variable of interest is either a dummy, which is "1" if oil was discovered in that county during the specific period or a variable indicating the number of years passed since the discovery of oil. The former regressions give the effect of the discovery of oil over a period of at most 30 years, while the latter regressions, which use only the counties where oil had been found, test whether there are any longer-term changes caused by oil extraction. Standard errors are always adjusted for clustering on the county level to avoid the well-known Moulton-problem (Moulton, 1990).

There are two broad concerns with this estimation approach. The first is that counties with and without oil may be in some way different from each other. A quick look at Table 2 that plots differences in 1880 between individuals living in countries without oil and individuals living in counties where oil was found between 1880 and 1910 reveals that there are indeed differences in certain characteristics. In short, counties where oil was discovered in later years seem to have leaned more towards agriculture and also have slightly lower human capital.

(Table 2 about here.)

An important question is whether these differences can be expected to matter for the results. Given the presence of the county specific fixed effects $\eta_c$, identification of $\tau$

[7] The first state to introduce compulsory schooling laws was Massachusetts in 1852, the last Alaska in 1929 preceded by Mississippi in 1918.

uses variation within counties over time. Base differences in characteristics between counties are relatively harmless in this context.

However, the second, more serious potential problem with this difference-in-differences-type strategy is that discoveries of oil in a given time period might be correlated with differences in unobserved trends. To overcome this problem I rely on a second identification strategy: I exploit the fact that the discovery of oil during the observation period of this paper spread out towards the South West and North East of the USA from Venango County, Pennsylvania, where the first oil well was drilled in Oil Creek. This fact suggests that the distance of a specific county from Venango County is a predictor of whether oil was found in that county in a given year.[8] As the distance from Venango County is uninfluenced by any unobserved county-level trends, it is a suitable instrument for $D_i$ in equation (1). An area of concern, however, is that counties farther away from Venango County might differ in baseline outcomes from counties closer to Venango County. To test whether this is the case, Table 3 shows results of a regression of distance to Venango County (in 100km), the control variables used in every regression and state fixed effects on the respective outcome using the 1880 cross section. As oil had already been discovered in some countries, I present results based on two specifications including and excluding controls for oil production in the respective county in 1880. If the instrument is valid, distance should have a small and insignificant effect on any of the outcomes.

---

[8] The distance is calculated as the ellipsoidal distance between two counties' centroids using the ado-file *geodist* by Robert Picar, which uses the formulas derived by Vincenty (1975).

(Table 3 about here.)

The results generally show no significant and large (within state) differences between counties closer and farther away from Venango County. Exceptions are the two measures of education where counties closer to Venango County appear to have somewhat more favorable characteristics. However, the differences observed here are generally small: Being 100km closer to Venango County changes the measures of education by about half a percentage point. Combined with a fairly strong instrument with first stage F-values for the excluded instruments between 10 and 32 as the following sections show, this finding attenuates concerns regarding the validity of the instrument. Furthermore, the evidence from both estimation strategies is generally qualitatively similar, although the IV estimates are sometimes insignificant, which can be explained by the lower efficiency of 2SLS.

## IV. Results

### *A. Urbanization*

I begin the analysis by looking at urbanization. Evidence by Michaels (2011) suggests that there were positive net migration flows into oil abundant counties by 1940, which he explains by higher income levels in these counties. This evidence is also in line with the anecdotal evidence on rapid urbanization following the discovery of oil quoted in Section II. Table 4a presents estimates based on equation (1); the 2SLS estimates as described above are shown in Table 4b.

(Tables 4a and 4b about here.)

The evidence in Table 4a suggests that the discovery of oil indeed moves around 4% of the population into urban areas. At the same time, the share of the population living on farms declines by a similar amount. The point estimates are relatively similar for men and women. We also see similar effects for the time passed since the discovery of oil in the oil producing counties, which appear to be slightly stronger for men. A possible explanation for this finding is that men migrated primarily for work reasons immediately after the discovery of oil to find work, while women followed their men only occasionally or with some delay, thus leading to somewhat weaker effects for them. Given that many cities grew and declined with the production of oil, such behavior would be rational if women migrated primarily for family reasons and only after they saw a chance to stay permanently.

The evidence from the IV estimates in Table 4b suggests a similar pattern of results. Note first that the first stage results indicate the expected negative relationship between a county's distance from Venango County and both the likelihood that oil was discovered and the years passed since its discovery. The instrument is also reasonable strong with first stage F-values between 12 and 32. Point estimates in this table generally support the evidence from Table 3a, although the estimates are often insignificant. However, this insignificance can be explained by the much lower efficiency of 2SLS relative to OLS.

***B. Labor market outcomes***

A core mechanism that may lead to the resource curse is a change in prices caused by the discovery of the natural resource, which makes working on the extraction of that resource attractive for many individuals and draws them from other sectors of the economy. In the following I consider whether the discovery of oil indeed drew people into the mining industry (or the labor force), whether it crowded out agriculture or manufacturing and whether it changed the economic situation of individuals. Table 5a and 5b present the results.

(Tables 5a and 5b about here.)

Starting again with the regression results in Table 5a, we see clear evidence that individuals, specifically men, were indeed drawn into the mining sector. The point estimates indicate that men were 3% more likely to work in mining in counties with oil. This effect is large relatively to the employment shares of the mining industry in both 1880 and 1910, when it employed between 2 and 3% of the male workforce. We also see a large increase in the share of men working in manufacturing, which runs contrary to the common perception that resource extraction crowds out manufacturing. In fact, the sector that is being crowded out in this case seems to be agriculture, while manufacturing seems to benefit from the discovery of oil. This result is similar to the findings by Michaels (2011) for 1940, suggesting that these long-term differences had their roots in the immediate aftermath of the discovery of oil. It also fits earlier evidence by Wright (1990) who highlights the importance of resource abundance for American industrial development. The results are very similar when looking at the years passed since the first discovery of oil. For women, we observe some effects for the likelihood of working in manufacturing, which increases by 5% respectively.

Looking at the IV results in Table 5b again yields relatively similar results for manufacturing and agriculture although the results are again insignificant. The coefficients in the regressions for being in the mining sector, however, have opposite signs and are sometimes insignificant. While the former observation is not surprising, the latter requires some explanation. One potential reason for this counterintuitive result could be reverse causality affecting the results in Table 5a. In this case, it would be increases in mining personnel, e.g., exploration teams, causing the discovery of oil and not vice versa. It should be noted, however, that this result does not fit either the historical evidence or the evidence in Michaels (2011) very well, which leaves the possibility that it is just a statistical artifact. The results for women show no real effect on the occupational structure.

Looking at labor market participation and proxies for the economic condition show some evidence that women were drawn into the labor force and that their economic situation improved when looking at the OLS estimates, while no real effect shown up in the IV estimates. For men, the evidence in Table 5a suggests (a) relative minor effects on labor market participation, which is not surprising as labor market participation rates were very high to begin with and (b) improvements in the economic situation as measured by the occupational income score. The latter changes are also rather large relative to the 1880 base value, amounting to increases of 3.6% for men and of 6.5% for women. The IV-estimates in Table 5b again broadly confirm these results, although their lower efficiency leads to slightly more erratic estimates and quite often to insignificance of the results.

### *C. Education*

Natural resources and the specialization on their extraction are often thought to decrease incentives for education (see, e.g., Leamer et al., 1999; Gylfason, 2001). The common explanation for this finding is that (low-skilled) resource extraction tends to crowd out (higher-skilled) manufacturing. However, the results from the previous section as well as the long-term evidence by Michaels (2011) suggest that oil extraction in the US crowded out (low-skilled) agriculture. Furthermore, the evidence from the previous section also suggests that the employment share of manufacturing tended to increase after the discovery of oil. Both of these factors suggest that we might see an increase rather than a decrease in human capital accumulation following the discovery of oil.

(Tables 6a and 6b about here.)

Table 6a and 6b present the accompanying evidence. The evidence suggests some positive effects of oil discoveries on human capital. The OLS estimates indicate that both school attendance and literacy tend to increase with the duration passed since the discovery of oil. These results are broadly consistent with higher incentives to invest in human capital due to an increasing specialization in oil extraction and manufacturing and decreasing employment opportunities in agriculture. The IV results in Table 6b show (a) an increase in school attendance following the discovery of oil and (b) on balance a decrease in illiteracy both for the discovery of oil and the duration passed since that discovery.

Overall, it seems fair to say that the evidence is in favor of a positive effect on human capital accumulation for both men and women, which is consistent with an increase in employment opportunities for higher skilled workers in resource extraction and manufacturing. This finding is also in line with the results by Michaels (2011) who finds higher human capital in oil abundant counties in 1940. For women, the results are somewhat more mixed with one estimate indicating a negative effect on literacy. This finding mirrors the differences found for labor market outcomes in the preceding section: If changes in employment opportunities are the main driver of the observed effects, we would expect these to be smaller for women due to their much lower labor market participation.

## V. Conclusion

This paper investigated the economic effects of the US oil boom from 1880 to 1910, in particular several predictions arising from the literature on the resource curse. The substantial findings are that oil discoveries (a) increased urbanization at the expense of individuals living on farms, (b) led to sectoral shifts by drawing more (primarily male) workers into mining and manufacturing and away from agriculture, (c) improved the economic situation of men and women by drawing them into occupations that enjoyed a higher socio-economic status, and (d) broadly improved the education of men and women, which is likely related to the shifts in employment. The effects are generally somewhat weaker for women, which supports the idea that changes in employment opportunities were the main driver of these changes.

In terms of the resource curse, these results suggest that resource extraction does not carry a curse per se, but rather that circumstances matter. Compared to today's oil abundant states, the US were lucky that manufacturing and oil extraction appear to have been complements rather than substitutes during the period from 1880 to 1910, as many industries relied on petroleum and transport costs were still relatively high. Furthermore, one can also speculate that the political institutions in the US – being a democracy for roughly a century as well as a free market economy at that time – were better suited to deal with the consequences of resource discoveries that many of the autocratic countries that are nowadays associated with being "cursed" by oil. It is, however, important to stress that even in the US resource abundance carried a long term curse as oil abundant counties had no long-term advantage over counties without oil (Michaels, 2011) and grew slower at the end of the 20th century (Papyrakis and Gerlagh, 2007).

## REFERENCES

Aragón, Fernando and Juan Pablo Rub (2011): "Natural Resources and Local Economic Development: Evidence from a Peruvian Gold Mine", mimeo.

Bacon, Raymond Foss and William Allen Hamor (1916): "*The American Petroleum Industry, Volume I*", New York: McGraw-Hill.

Becker, Sascha O. and Ludger Woessmann (2009): "Was Weber wrong? A human capital theory of Protestant economic history", *Quarterly Journal of Economics 124 (2)*, pp. 531-596.

Brunnschweiler, Christa N. and Erwin H. Bulte, E. (2009): "Natural Resources and Violent Conflict: Resource Abundance, Dependence and the Onset of Civil Wars", *Oxford Economic Papers 61(4)*, pp. 651-674.

Caselli, Francesco and Coleman II, Wilbur John (2006): "On the theory of ethnic conflict", CEP Discussion Paper 732.

Caselli, Franceso and Guy Michaels (2009): "Do Oil Windfalls Improve Living Standards? Evidence from Brazil", *NBER Working Paper 15550*.

Collier, Paul and Goderis, Benedikt (2009): "Commodity Prices, Growth, and the Natural Resource Curse: Reconciling a Conundrum", *mimeo*, Oxford.

Corden, Max W. and Neary, Peter J. (1982): "Booming Sector and De-Industrialisation in a Small Open Economy." *Economic Journal 92(368)*, pp. 825-848.

Davidson, Paul (1963): "Public Policy Problems of the Domestic Crude Oil Industry" *American Economic Review 53(1)*, pp. 85-108.

Energy Information Administration (2009): "*Oil and Gas Field Code Master List 2008*", Office of Oil and Gas, U.S. Department of Energy, Washington, DC.

Engerman, Stanley L. and Sokoloff, Kenneth L. (1997): "Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies." in S.H. Haber (ed.), *How Latin America Fell Behind*. California: Stanford University Press.

Gylfason, Thorvaldur (2001): "Natural Resources, Education, and Economic Development," *European Economic Review 45(4-6)*, 847-859

Humphreys, Macartan, Jeffrey D. Sachs and Joseph E. Stiglitz (2007): "*Escaping The Resource Curse*". New York, NY: Columbia University Press.

Isham, Jonathan, Michael Woolcock, Lant Pritchett and Gwen Busby (2005). 'The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth', *World Bank Economic Review 19(2)*, pp. 141-174.

James, Alexander G. and Robert G. James (2011): "Do resource dependent regrion grow slower than they should?", *Economics Letters 111(3)*, pp. 194-196

Kolstad, Ivar (2009): "The Resource Curse: Which Institutions Matter?", *Applied Economics Letters 16(4)*, pp. 439-442.

Leamer, Edward E. Hugo Maul, Sergio Rodriquez and Peter K. Schott (1999): “Does natural resource abundance increase Latin American income inequality?", *Journal of Development Economics 59(1)*, pp. 3-42.

Leite, Carlos and Jens Weidmann (1999): “Does Mother Nature Corrupt? Natural Resources, Corruption and Economic Growth”, *IMF Working Papers 99/85*.

Libecap, Gary D. (1989): “The Political Economy of Crude Oil Cartelization in the United States, 1933-1972”, *Journal of Economic History 49(4)*, pp. 833-855.

Libecap, Gary D. and Wiggins, Steven N. (1984): “Contractual Responses to the Common Pool: Prorationing of Crude Oil Production”, *American Economic Review 74(1)*, pp. 87-98.

Mehlum, Halvor, Karl O. Moene and Ragnar Torvik (2006a): “Institutions and the Resource Curse”, *Economic Journal 116(508)*, pp.1–20

Mehlum, Halvor, Karl O. Moene and Ragnar Torvik (2006b): “Cursed by Resources or Institutions?”, *The World Economy 29(8)*, pp. 1117-1131.

Michaels, Guy (2011): “The Long Term Consequences of Resource-Based Specialization”, *Economic Journal 121(551)*, pp. 31-57.

Moulton, Brent R. (1990): “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units”, *Review of Economic and Statistics 72(2)*, pp. 334-338.

Naritomi, Joana, Rodrigo R. Soares, and Juliano J. Assunção (2007): “Rent Seeking and the Unveiling of 'De Facto' Institutions: Development and Colonial Heritage within Brazil,” *NBER Working Paper No. 13545*.

Owen, Edgar Wesley (1975): “*Trek of the Oil Finders: A History of Exploration for Petroleum*”, Tulsa, Oklahoma: The American Association of Petroleum Geologists.

Papyrakis, Elissaios and Reyer Gerlagh (2007): “Resource abundance and economic growth in the United States”, *European Economic Review 51(4)*, pp. 1011–1039.

van der Ploeg, Frederick (2011): “Natural Resources: Curse or Blessing?”, *Journal of Economic Literature 49(2)*, pp. 366-420.

Ross, Michael L. (1999), “The Political Economy of the Resource Curse”, *World Politics 51(2)*, pp. 297-322.

Ross, Michael L. (2006): “A Closer Look at Oil, Diamonds, and Civil War,” *Annual Review of Political Science 9*, pp. 265–300.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder and Matthew Sobek (2010): “*Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*”, Minneapolis, MN: Minnesota Population Center [producer and distributor].

Sachs, Jeffrey D. and Warner, Andrew M. (1995): “Natural Resource Abundance and Economic Growth”, *NBER Working Paper 5398.*

Sachs, Jeffrey D. and Warner, Andrew M. (2001): “The Curse of Natural Resources”, *European Economic Review 45(4-6)*, pp. 827-838.

Tornell, Aaron and Philip R. Lane(1999) “The Voracity Effect”, *American Economic Review 89(1)*, pp. 22-46

Vicente, Pedro C. (2010): “Does Oil Corrupt? Evidence from a Natural Experiment in West Africa”, *Journal of Development Economics 92(1)*, pp. 28-38.

Vincenty, Thaddeus (1975): “Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations”, *Survey Review 22(176)*, pp. 88-93.

Wright, Gavin (1986): “*Old South, New South*”, New York: Basic Books.

Wright, Gavin (1990): “The Origins of American Industrial Success, 1879-1940”, *American Economic Review 80(4)*, pp. 651-668

**Table 1:**
*Descriptive statistics*

| | Pooled sample | | | 1880 | | | 1910 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. |
| **Men** | | | | | | | | | |
| Oil in county of residence in respective year (1=yes) | 639769 | 0.134 | 0.340 | 219969 | 0.059 | 0.235 | 419800 | 0.173 | 0.378 |
| Years passed since discovery of oil (only counties with oil) | 85577 | 21.752 | 14.284 | 12869 | 11.257 | 7.158 | 72708 | 23.610 | 14.426 |
| Age (years) | 639769 | 26.342 | 19.012 | 219969 | 24.637 | 18.739 | 419800 | 27.235 | 19.093 |
| First generation migrant | 639769 | 0.172 | 0.377 | 219969 | 0.159 | 0.366 | 419800 | 0.178 | 0.383 |
| Second generation migrant) | 639769 | 0.221 | 0.415 | 219969 | 0.204 | 0.403 | 419800 | 0.229 | 0.420 |
| Lives in urban area (1=yes) | 639769 | 0.399 | 0.490 | 219969 | 0.274 | 0.446 | 419800 | 0.464 | 0.499 |
| Lives on farm (1=yes) | 639769 | 0.363 | 0.481 | 219969 | 0.455 | 0.498 | 419800 | 0.314 | 0.464 |
| Currently attending school (1=yes) | 639769 | 0.207 | 0.405 | 219969 | 0.184 | 0.387 | 419800 | 0.220 | 0.414 |
| Illiterate (1=yes) | 639769 | 0.051 | 0.221 | 219969 | 0.066 | 0.248 | 419800 | 0.044 | 0.205 |
| In the labor force (1=yes) | 416714 | 0.905 | 0.293 | 134140 | 0.909 | 0.288 | 282574 | 0.904 | 0.295 |
| Miner (1=yes) | 377323 | 0.027 | 0.163 | 121907 | 0.018 | 0.132 | 255416 | 0.032 | 0.175 |
| Manufacturing worker (1=yes) | 377323 | 0.427 | 0.495 | 121907 | 0.378 | 0.485 | 255416 | 0.450 | 0.498 |
| Farmer, agricultural workers (1=yes) | 377323 | 0.349 | 0.477 | 121907 | 0.456 | 0.498 | 255416 | 0.298 | 0.458 |
| Occupational income score | 377323 | 22.045 | 10.986 | 121907 | 20.431 | 10.617 | 255416 | 22.816 | 11.075 |
| **Women** | | | | | | | | | |
| Oil in county of residence in respective year (1=yes) | 606690 | 0.132 | 0.338 | 211450 | 0.058 | 0.234 | 395240 | 0.171 | 0.377 |
| Years passed since discovery of oil (only counties with oil) | 79890 | 21.580 | 14.163 | 12259 | 11.348 | 7.271 | 67631 | 23.434 | 14.316 |
| Age (years) | 606690 | 25.975 | 19.130 | 211450 | 24.409 | 18.720 | 395240 | 26.813 | 19.293 |
| First generation migrant | 606690 | 0.146 | 0.353 | 211450 | 0.143 | 0.350 | 395240 | 0.148 | 0.355 |
| Second generation migrant) | 606690 | 0.232 | 0.422 | 211450 | 0.210 | 0.407 | 395240 | 0.244 | 0.429 |
| Lives in urban area (1=yes) | 606690 | 0.416 | 0.493 | 211450 | 0.292 | 0.455 | 395240 | 0.482 | 0.500 |
| Lives on farm (1=yes) | 606690 | 0.345 | 0.475 | 211450 | 0.427 | 0.495 | 395240 | 0.301 | 0.459 |
| Currently attending school (1=yes) | 606690 | 0.214 | 0.410 | 211450 | 0.179 | 0.383 | 395240 | 0.232 | 0.422 |
| Illiterate (1=yes) | 606690 | 0.054 | 0.227 | 211450 | 0.077 | 0.267 | 395240 | 0.042 | 0.201 |
| In the labor force (1=yes) | 389865 | 0.188 | 0.391 | 128541 | 0.127 | 0.333 | 261324 | 0.218 | 0.413 |
| Miner (1=yes) | 73413 | 0.000 | 0.020 | 16374 | 0.000 | 0.000 | 57039 | 0.001 | 0.023 |
| Manufacturing worker (1=yes) | 73413 | 0.317 | 0.465 | 16374 | 0.367 | 0.482 | 57039 | 0.302 | 0.459 |
| Farmer, agricultural workers (1=yes) | 73413 | 0.094 | 0.292 | 16374 | 0.069 | 0.253 | 57039 | 0.101 | 0.302 |
| Occupational income score | 73413 | 16.739 | 9.309 | 16374 | 15.012 | 9.348 | 57039 | 17.235 | 9.239 |

**Table 2:**

*Comparison of county characteristics before discovery of oil, 1880*

| | **Never oil** | **Oil discovered 1880 to 1910** |
|---|---|---|
| Male | 0.510 | 0.511 |
| | (0.500) | (0.500) |
| Age (years) | 24.582 | 24.192 |
| | (18.733) | (18.692) |
| First generation migrant | 0.160 | 0.097 |
| | (0.367) | (0.296) |
| Second generation migrant) | 0.209 | 0.177 |
| | (0.407) | (0.382) |
| Lives in urban area (1=yes) | 0.295 | 0.199 |
| | (0.456) | (0.399) |
| Lives on farm (1=yes) | 0.434 | 0.512 |
| | (0.496) | (0.500) |
| Currently attending school (1=yes) | 0.176 | 0.205 |
| | (0.381) | (0.404) |
| Illiterate (1=yes) | 0.074 | 0.064 |
| | (0.262) | (0.244) |
| In the labor force (1=yes) | 0.530 | 0.506 |
| | (0.499) | (0.500) |
| Miner (1=yes) | 0.015 | 0.016 |
| | (0.120) | (0.125) |
| Manufacturing worker (1=yes) | 0.382 | 0.318 |
| | (0.486) | (0.466) |
| Farmer, agricultural workers (1=yes) | 0.400 | 0.498 |
| | (0.490) | (0.500) |
| Occupational income score | 19.873 | 18.942 |
| | (10.611) | (10.710) |
| Observations | 354,842 | 51,449 |

Note: Means, standard deviations in parentheses. All data are from the 1880 census. Number of cases differ between variables, number given is total number of observations.

**Table 3:**
*Distance to Venango County, PA, and county characteristics in 1880*

| | Lives in urban area | Lives on farm | Attends school (younger than 14) | Illiterate (older than 10) | In the labour force | Miner | Manufacturing worker | Agricultural worker | Occupational income score |
|---|---|---|---|---|---|---|---|---|---|
| **Men** | | | | | | | | | |
| Not controlling for oil in 1880 | | | | | | | | | |
| Distance to Venango County, PA(in 100km) | 0.0074 | -0.0032 | -0.0047*** | 0.0035** | -0.0002 | -0.0010 | -0.0005 | -0.0049 | 0.0936 |
| | (0.0182) | (0.0102) | (0.0016) | (0.0014) | (0.0009) | (0.0020) | (0.0053) | (0.0108) | (0.1926) |
| Controlling for oil in 1880 | | | | | | | | | |
| Distance to Venango County, PA(in 100km) | 0.0009 | 0.0007 | -0.0046*** | 0.0039*** | -0.0002 | -0.0008 | -0.0030 | -0.0008 | 0.0345 |
| | (0.0168) | (0.0107) | (0.0017) | (0.0015) | (0.0010) | (0.0019) | (0.0058) | (0.0113) | (0.1927) |
| N | 219969 | 219969 | 219969 | 219969 | 134140 | 121907 | 121907 | 121907 | 121907 |
| **Women** | | | | | | | | | |
| Not controlling for oil in 1880 | | | | | | | | | |
| Distance to Venango County, PA(in 100km) | 0.0103 | -0.0032 | -0.0040*** | 0.0040** | 0.0046 | 0.0000 | 0.0126 | 0.0026 | 0.3767*** |
| | (0.0189) | (0.0101) | (0.0014) | (0.0018) | (0.0039) | (.) | (0.0081) | (0.0029) | (0.1457) |
| Controlling for oil in 1880 | | | | | | | | | |
| Distance to Venango County, PA(in 100km) | 0.0030 | 0.0011 | -0.0028* | 0.0040** | 0.0013 | 0.0000 | 0.0046 | 0.0045 | 0.3798*** |
| | (0.0173) | (0.0105) | (0.0015) | (0.0019) | (0.0033) | (.) | (0.0074) | (0.0035) | (0.1386) |
| N | 211450 | 211450 | 211450 | 211450 | 128541 | 16374 | 16374 | 16374 | 16374 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level. All estimates control for age and age squared, first and second generation migration status and state fixed effects.

**Table 4a:**
*Oil and urbanization, OLS estimates*

| | Men | | Women | |
|---|---|---|---|---|
| | **Lives in urban area** | **Lives on farm** | **Lives in urban area** | **Lives on farm** |
| All counties | | | | |
| Oil found (1 = yes) | 0.0400*** | -0.0549*** | 0.0439*** | -0.0469*** |
| | (0.0140) | (0.0116) | (0.0152) | (0.0116) |
| Observations | 638587 | 638587 | 605692 | 605692 |
| $R^2$ | 0.5343 | 0.3379 | 0.5384 | 0.3291 |
| Only counties with oil in respective years | | | | |
| Years since first oil found | 0.0126*** | -0.0028*** | 0.0047*** | -0.0056*** |
| | (0.0001) | (0.0001) | (0.0000) | (0.0000) |
| Observations | 85334 | 85334 | 79695 | 79695 |
| $R^2$ | 0.3426 | 0.2372 | 0.3545 | 0.2283 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level. All estimates control for age and age squared, first and second generation migration status, state-year-fixed effects and county fixed effects.

**Table 4b:**
*Oil and urbanization, IV-estimates*

| | **Men** | | **Women** | |
|---|---|---|---|---|
| | Lives in urban area | Lives on farm | Lives in urban area | Lives on farm |
| | All counties | | | |
| | **First stage, dep. var. oil found (1 = yes)** | | | |
| Distance to Venango County, PA (km)[a] | -0.0007*** | -0.0007*** | -0.0008*** | -0.0008*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| $R^2$ | 0.0919 | 0.0919 | 0.0952 | 0.0952 |
| F (excluded instruments) | 32.5092 | 32.5092 | 31.0268 | 31.0268 |
| | **Second stage** | | | |
| Oil found (1 = yes) | 0.0960 | -0.0403 | 0.0894 | -0.0474 |
| | (0.2061) | (0.1103) | (0.2022) | (0.1075) |
| N | 638587 | 638587 | 605692 | 605692 |
| | Only counties with oil in respective years | | | |
| | **First stage, dep. var. years since first oil found** | | | |
| Distance to Venango County, PA (km)[a] | -0.0300*** | -0.0300*** | -0.0315*** | -0.0315*** |
| | (0.0085) | (0.0085) | (0.0083) | (0.0083) |
| $R^2$ | 0.0621 | 0.0621 | 0.0680 | 0.0680 |
| F (excluded instruments) | 12.4319 | 12.4319 | 14.3953 | 14.3953 |
| | **Second stage** | | | |
| Years since first oil found | 0.0126* | -0.0053 | 0.0100 | -0.0039 |
| | (0.0075) | (0.0042) | (0.0065) | (0.0038) |
| N | 85334 | 85334 | 79695 | 79695 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level. All estimates control for age and age squared, first and second generation migration status and state-year-fixed effects.

[a] Computed as the ellipsoidal distance using equations by Vincenty (1975) and the WGS 1984 datum.

**Table 5a:**

*Oil and labor market outcomes, OLS estimates, only individuals older than 16 years*

| | **In the labour force** | **Miner** | **Manufacturing worker** | **Agricultural worker** | **Occupational income score** |
|---|---|---|---|---|---|
| | | | **Men** | | |
| | | | All counties | | |
| Oil found (1 = yes) | 0.0040 | 0.0293*** | 0.0472*** | -0.0498*** | 0.7534*** |
| | (0.0047) | (0.0074) | (0.0119) | (0.0127) | (0.2259) |
| Observations | 415765 | 376455 | 376455 | 376455 | 376455 |
| $R^2$ | 0.1460 | 0.2237 | 0.1932 | 0.3531 | 0.1826 |
| | | | Only counties with oil in respective years | | |
| Years since first oil found | -0.0011 | 0.0123*** | 0.0107*** | -0.0027*** | 0.0423*** |
| | (0.0007) | (0.0001) | (0.0001) | (0.0001) | (0.0040) |
| Observations | 56730 | 51259 | 51259 | 51259 | 51259 |
| $R^2$ | 0.1709 | 0.1846 | 0.1735 | 0.2597 | 0.1352 |
| | | | **Women** | | |
| | | | All counties | | |
| Oil found (1 = yes) | 0.0102* | 0.0003 | 0.0483** | -0.0037 | 1.0065*** |
| | (0.0057) | (0.0006) | (0.0196) | (0.0109) | (0.3692) |
| Observations | 389106 | 73114 | 73114 | 73114 | 73114 |
| $R^2$ | 0.1019 | 0.0890 | 0.1497 | 0.4826 | 0.1544 |
| | | | Only counties with oil in respective years | | |
| Years since first oil found | 0.0067*** | 0.0000 | 0.0025*** | -0.0012 | -0.0480 |
| | (0.0001) | (0.0000) | (0.0007) | (0.0053) | (0.0681) |
| Observations | 52127 | 8996 | 8996 | 8996 | 8996 |
| $R^2$ | 0.0669 | 0.0320 | 0.0893 | 0.3704 | 0.1146 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level. All estimates control for age and age squared, first and second generation migration status, state-year-fixed effects and county fixed effects.

**Table 5b:**

*Oil and labor market outcomes, IV estimates, only individuals older than 16 years*

| | Men | | | | | Women | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **In the labour force** | **Miner** | **Manufacturing worker** | **Agricultural worker** | **Occupational income score** | **In the labour force** | **Miner** | **Manufacturing worker** | **Agricultural worker** | **Occupational income score** |
| All counties | | | | | | | | | | |
| **First stage, dep. var. oil found (1 = yes)** | | | | | | | | | | |
| Distance to Venango County, PA (km)[a] | -0.0008*** | -0.0008*** | -0.0008*** | -0.0008*** | -0.0008*** | -0.0008*** | -0.0009*** | -0.0009*** | -0.0009*** | -0.0009*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| $R^2$ | 0.0947 | 0.0955 | 0.0955 | 0.0955 | 0.0955 | 0.1021 | 0.1245 | 0.1245 | 0.1245 | 0.1245 |
| F (excluded instruments) | 32.2857 | 31.8462 | 31.8462 | 31.8462 | 31.8462 | 30.8289 | 22.2374 | 22.2374 | 22.2374 | 22.2374 |
| **Second stage** | | | | | | | | | | |
| Oil found (1 = yes) | 0.0092 | -0.0296 | 0.1127** | -0.0652 | 1.0490 | 0.0043 | -0.0012 | -0.0574 | 0.0015 | -1.2299 |
| | (0.0085) | (0.0431) | (0.0536) | (0.1107) | (2.0550) | (0.0420) | (0.0010) | (0.0548) | (0.0300) | (1.0039) |
| N | 415765 | 376455 | 376455 | 376455 | 376455 | 389106 | 73114 | 73114 | 73114 | 73114 |
| Only counties with oil in respective years | | | | | | | | | | |
| **First stage, dep. var. years since first oil found** | | | | | | | | | | |
| Distance to Venango County, PA (km)[a] | -0.0294*** | -0.0298*** | -0.0298*** | -0.0298*** | -0.0298*** | -0.0307*** | -0.0318*** | -0.0318*** | -0.0318*** | -0.0318*** |
| | (0.0090) | (0.0091) | (0.0091) | (0.0091) | (0.0091) | (0.0087) | (0.0102) | (0.0102) | (0.0102) | (0.0102) |
| $R^2$ | 0.0591 | 0.0605 | 0.0605 | 0.0605 | 0.0605 | 0.0646 | 0.0716 | 0.0716 | 0.0716 | 0.0716 |
| F (excluded instruments) | 10.7652 | 10.7297 | 10.7297 | 10.7297 | 10.7297 | 12.4768 | 9.7010 | 9.7010 | 9.7010 | 9.7010 |
| **Second stage** | | | | | | | | | | |
| Years since first oil found | 0.0001 | -0.0042* | 0.0064* | -0.0093* | 0.1933* | 0.0011 | -0.0001 | -0.0021 | 0.0019 | -0.0120 |
| | (0.0008) | (0.0024) | (0.0034) | (0.0051) | (0.0991) | (0.0014) | (0.0001) | (0.0035) | (0.0024) | (0.0540) |
| N | 56730 | 51259 | 51259 | 51259 | 51259 | 52127 | 8996 | 8996 | 8996 | 8996 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level. All estimates control for age and age squared, first and second generation migration status and state-year-fixed effects.
[a] Computed as the ellipsoidal distance using equations by Vincenty (1975) and the WGS 1984 datum.

**Table 6a:**
*Oil and education, OLS estimates*

| | Men | | Women | |
|---|---|---|---|---|
| | **Attends school (younger than 14)** | **Illiterate (older than 10)** | **Attends school (younger than 14)** | **Illiterate (older than 10)** |
| All counties | | | | |
| Oil found (1 = yes) | -0.0003 | -0.0019 | -0.0034 | -0.0049 |
| | (0.0076) | (0.0053) | (0.0085) | (0.0065) |
| N | 211199 | 478593 | 205027 | 449996 |
| r2 | 0.5759 | 0.0958 | 0.5766 | 0.1314 |
| Only counties with oil in respective years | | | | |
| Years since first oil found | 0.0046*** | -0.0037* | 0.0037*** | 0.0019*** |
| | (0.0000) | (0.0020) | (0.0000) | (0.0000) |
| N | 27042 | 64930 | 25983 | 60067 |
| r2 | 0.6436 | 0.0919 | 0.6418 | 0.1084 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and+ 1% level. All estimates control for age and age squared, first and second generation migration status, state-year-fixed effects and county fixed effects.

**Table 6b:**

*Oil and education, IV estimates*

| | **Men** | | **Women** | |
|---|---|---|---|---|
| | **Attends school (younger than 14)** | **Illiterate (older than 10)** | **Attends school (younger than 14)** | **Illiterate (older than 10)** |
| All counties | | | | |
| **First stage, dep. var. oil found (1 = yes)** | | | | |
| Distance to Venango County, PA (km)[a] | -0.0007*** | -0.0008*** | -0.0007*** | -0.0008*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| $R^2$ | 0.0865 | 0.0939 | 0.0830 | 0.1001 |
| F (excluded instruments) | 32.6521 | 32.3330 | 31.2362 | 31.0066 |
| **Second stage** | | | | |
| Oil found (1 = yes) | 0.0261* | -0.0206 | 0.0208 | -0.0428** |
| | (0.0156) | (0.0179) | (0.0158) | (0.0198) |
| N | 211199 | 478593 | 205027 | 449996 |
| Only counties with oil in respective years | | | | |
| **First stage, dep. var. years since first oil found** | | | | |
| Distance to Venango County, PA (km)[a] | -0.0310*** | -0.0297*** | -0.0332*** | -0.0312*** |
| | (0.0078) | (0.0088) | (0.0077) | (0.0085) |
| $R^2$ | 0.0679 | 0.0605 | 0.0741 | 0.0667 |
| F (excluded instruments) | 15.9655 | 11.4712 | 18.3161 | 13.2984 |
| **Second stage** | | | | |
| Years since first oil found | 0.0004 | -0.0022* | 0.0009 | -0.0019* |
| | (0.0009) | (0.0013) | (0.0009) | (0.0011) |
| N | 27042 | 64930 | 25983 | 60067 |

Coefficients, standard errors adjusted for clustering on the county level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level. All estimates control for age and age squared, first and second generation migration status and state-year-fixed effects.

[a] Computed as the ellipsoidal distance using equations by Vincenty (1975) and the WGS 1984 datum.

**Figure 1:**
*Oil discoveries in the US, 1859 to 2008*

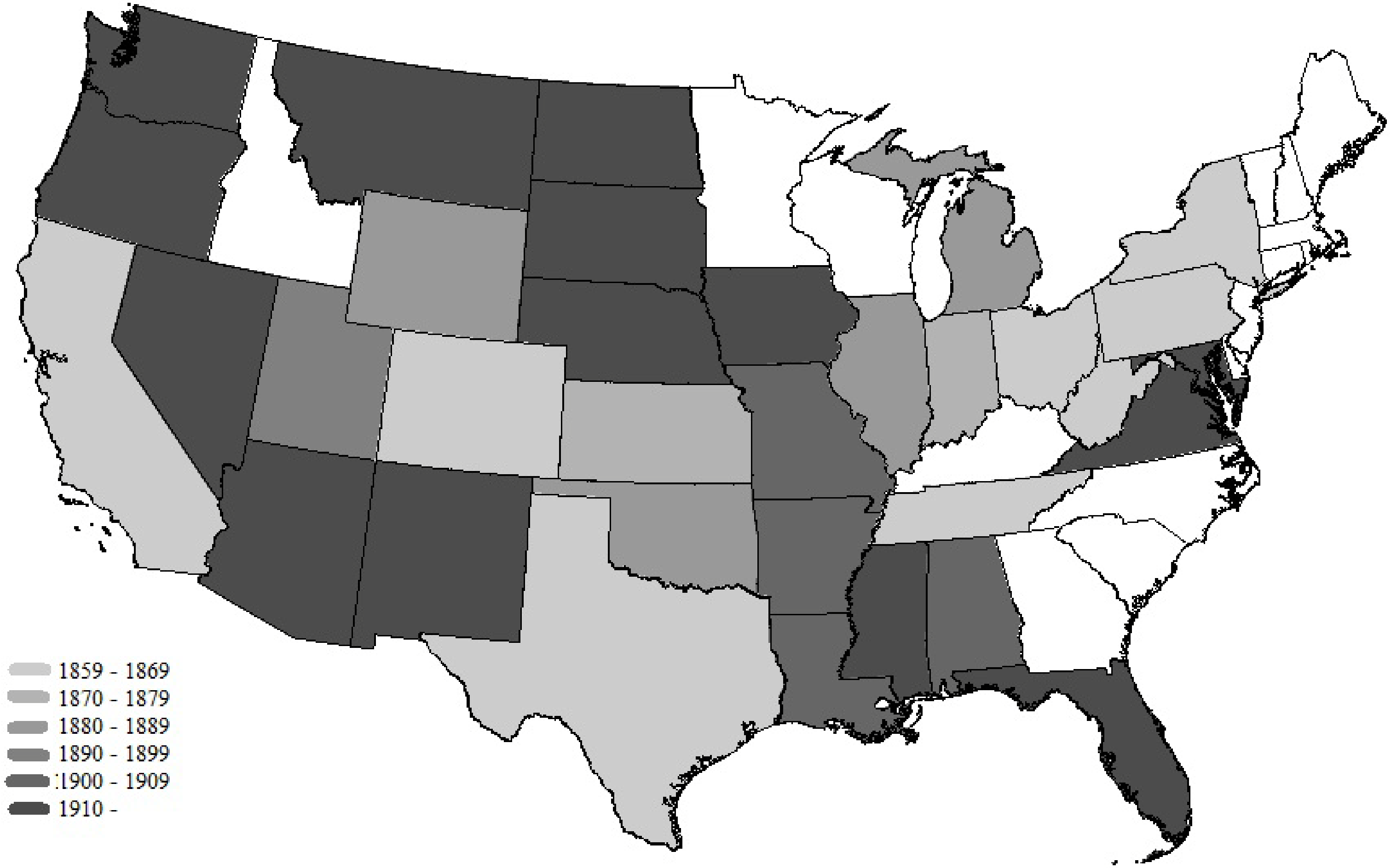
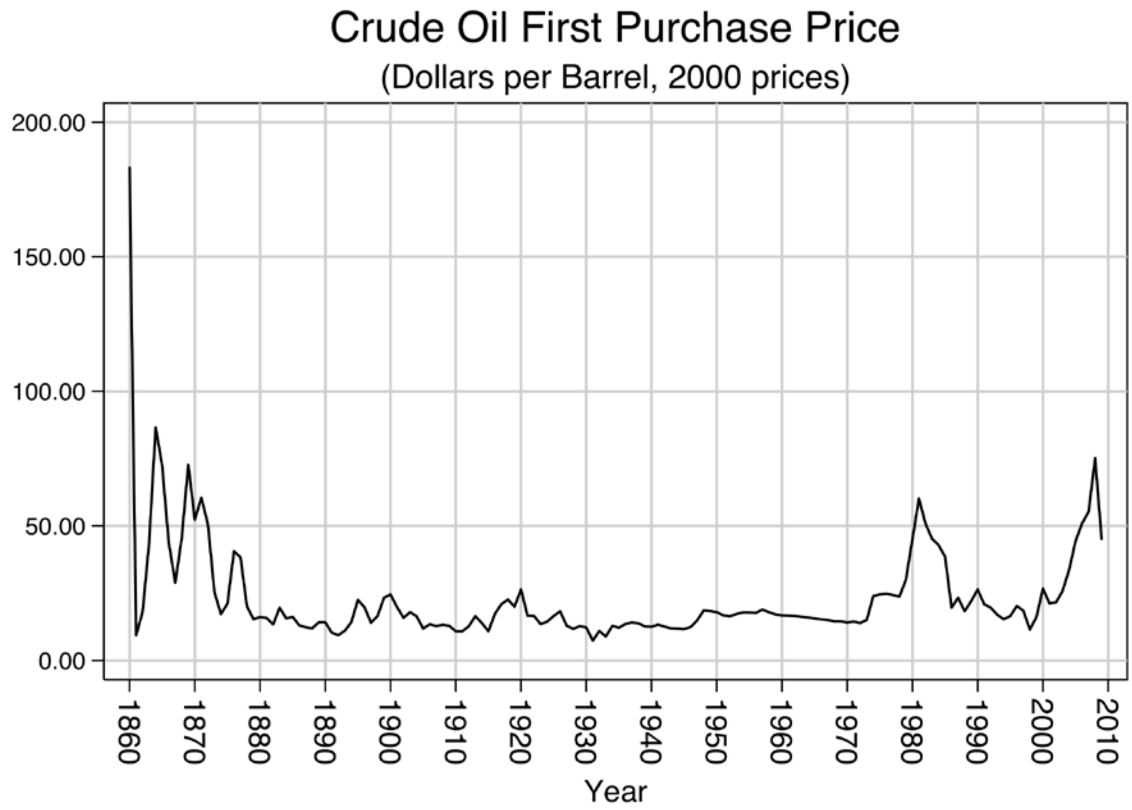
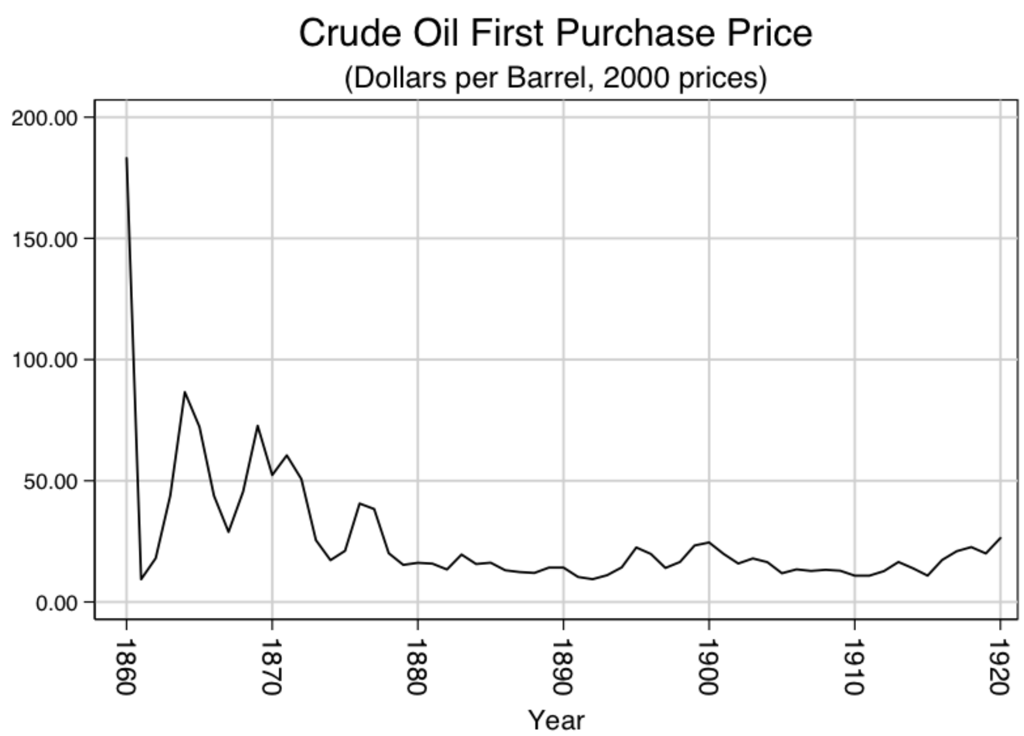


**Figure 2:**
*Crude Oil First Purchase Price, 2000 prices*
Panel (a) 1860-2010

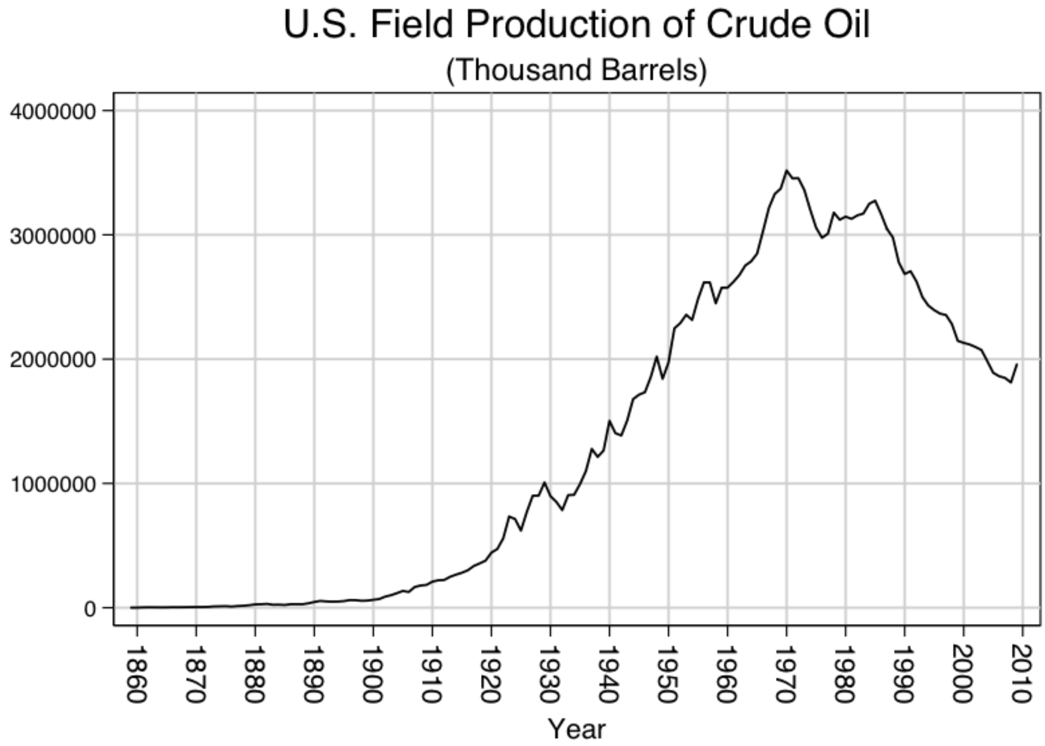



Panel (b) 1860 to 1920

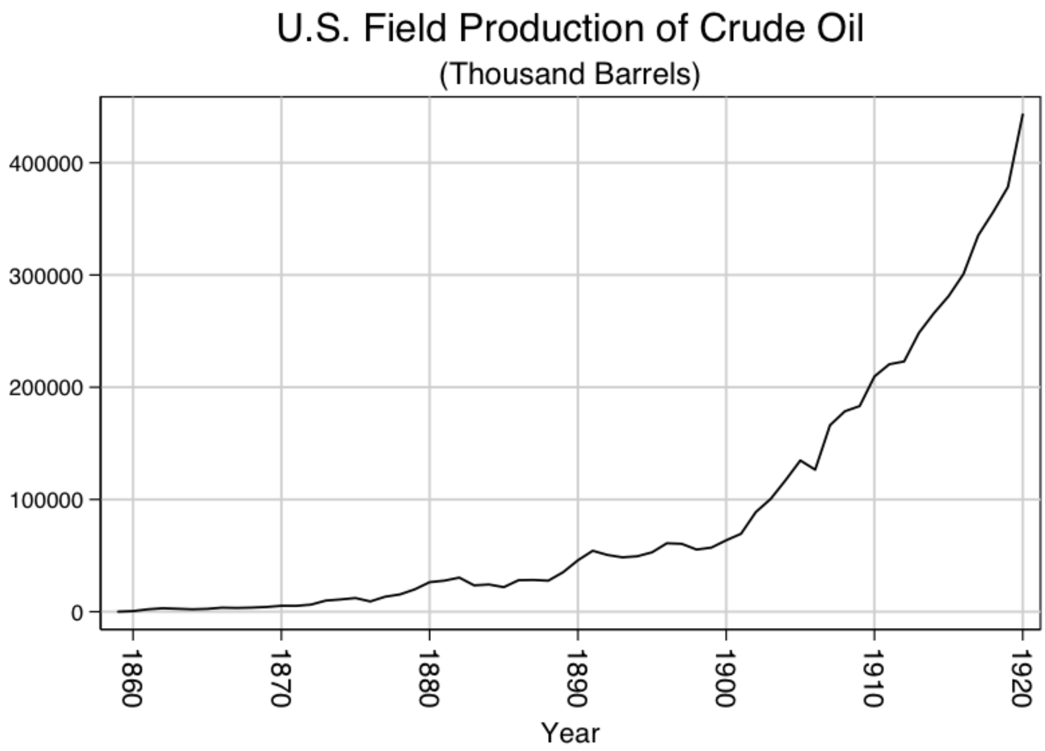



Source: Energy Information Administration

**Figure 3:**
*US Field Production of Crude Oil*
Panel (a): 1860 to 2010




Panel (b): 1860 to 1920




Source: Energy Information Administration