The Long-Run Effects of Parental Wealth Shocks on Children *

Cache Ellsworth† Ian Fillmore‡ Adrian Haws§ Joseph Price¶

December 16, 2023

Abstract

What are the causal effects of parental wealth on children's outcomes? Beginning with the famous land run of 1889, initial homesteaders in Oklahoma Territory raced to claim plots of land unaware that oil lay hidden beneath their feet. We link initial homesteaders to the locations of oil discoveries and develop new methods to link them to their children in the 1940 census, which allows us to examine the impacts of parental wealth shocks on children's wealth, income, labor supply, education, and migration.

Keywords: Intergenerational mobility, Intergenerational transfer of wealth

JEL Codes: J62, N32

*This research is supported by grants from the Weidenbaum Center at Washington University in St. Louis.

†University of Wisconsin-Madison, crellsworth@wisc.edu.

‡Washington University in St. Louis, ianfillmore@wustl.edu.

§Cornell University, aah226@cornell.edu.

¶Brigham Young University, joseph_price@byu.edu.

1 Introduction

Several recent papers have estimated rates of intergenerational economic mobility in the United States across time periods (Olivetti and Paserman (2015), Feigenbaum (2018), Song et al. (2019), Buckles et al. (2023b), Ward (2023)) and regions (Chetty et al. (2014), Connor and Storper (2020)). Differences in mobility rates suggest an important role for national and local policies in shaping intergenerational mobility. However, nearly all of these papers estimate correlations, not the causal effect of parental wealth or income on children. We exploit the quasi-random nature of oil discoveries in early twentieth century Oklahoma to estimate the causal long-run impact of a large wealth shock on children's outcomes, including direct measures of income, wealth, and education.

Identifying the causal effect of parental wealth on child outcomes is challenging because children may also be impacted by unobserved parental traits that are correlated with wealth, such as health, education, social capital, and cultural values. We combine data on the locations and drill dates of all oil wells drilled in Oklahoma with homesteader land records to construct precise measures of the dates of oil discovery for each homesteader. Linking homesteader records to the census presents an additional challenge because the homesteader records only recorded the homesteader's name. We develop a new linking method that exploits the detailed geographic information in the homesteader records, especially information about neighbors. Our method achieves a surprisingly high match rate and may be useful in other contexts where researchers need to link records with sparse demographic data but rich geographic information to the census. We then use cross-census links from the Census Tree project (Buckles et al., 2023a) to follow children into adulthood. After verifying that the presence of oil is unrelated to the characteristics of homesteading families, we estimate the causal effect of oil discoveries by comparing the children of homesteaders who were fortunate enough to have oil with their neighbors who were not as lucky.

We find that, relative to neighboring children, children of homesteaders with oil present are two percentage points more likely to own a home and one percentage point more likely to have a home value over $5,000 in the 1940 census. Additionally, they receive non-wage income (and pursue occupations outside of farming) at a two percentage point higher rate. We find no evidence of effects on wage income, hours worked, or labor force participation for the full sample, but we find a treatment effect of two fewer hours worked among children who were younger when oil was discovered. Although wealth does not appear affect education, children of homesteaders with oil are three percentage points more likely to be living in a city and this treatment effect increases to six percentage points for children who were younger when oil was discovered.

A small number of papers have used natural experiments to estimate the causal effects of parental wealth shocks on child outcomes. Bulman et al. (2021) find modest positive effects of winning the lottery on children's college attendance. Bleakley and Ferrie (2016) leverage an

1832 land lottery in Georgia and find that winning the lottery—a wealth shock roughly the size of median wealth at the time—increased fertility but did not increase children's wealth, occupational status, or literacy 50 years later. Ager et al. (2021) find that, after emancipation, children of former slaveholders obtained similar occupational status as the children of wealthy parents who had not owned slaves. Our paper differs from these in several respects. First, we employ cutting-edge linking methods to obtain a much larger sample. Second, the 1940 census is the first census to directly observe educational attainment and income for the full population, allowing us to gain insight into these important outcomes. Third, while the Georgia Land Lottery and emancipation occurred within the context of a nineteenth century agrarian society, the Oklahoma oil discoveries occurred within the context of a twentieth century industrialized economy. Fourth, oil shocks themselves may be a more "pure" form of wealth shock than winning land in a lottery since the value of land depends on the lottery winner's skills for farming, whereas oil royalties represent a genuinely passive stream of income. In Oklahoma, homesteaders could even sell the land while retaining mineral rights for themselves and, in many cases, passing those rights on to their children.

2 Background

The settlement of Oklahoma has a unique history. Modern day Oklahoma was originally called Indian Territory and was designated as a place for the resettling of American Indian tribes. However, beginning with the first land run of 1889, Indian Territory was gradually opened up to outside settlers, eventually resulting in the formation of Oklahoma Territory in the west and Indian Territory in the east. The land run of 1889 was little more than an organized race with each plot of land awarded to the first settler to claim it. Additional areas were opened up through land runs and land lotteries over the next several years. In 1901, the last major land opening occurred in modern-day southwest Oklahoma. Six years later, in 1907, Oklahoma Territory and Indian Territory were admitted into the United States as the state of Oklahoma. Thus, settlement occurred quickly within a relatively narrow time frame.

The first major oil discovery in modern day Oklahoma occurred in 1897 near Bartlesville in Indian Territory. Beginning in 1901, a series of oil discoveries further south transformed Tulsa into the "Oil Capital of the World." At the time of statehood in 1907, Oklahoma was the nation's largest oil producer, yet only a fraction of its true potential had yet been discovered. Over the next few decades, discoveries continued further south and west bringing a windfall to those lucky enough to own land with oil.

Oil wells were financed by "wildcatters," prospectors who drilled exploratory wells throughout the territory after signing an oil lease with the land owners. Land owners received royalties from the sale of oil and gas, typically at the rate of 12.5% of production value. Using a back-of-the-envelope calculation based on historical oil prices and the production amounts of the

Tonkawa-Three Sands oil field, land owners stood to gain between $160,000 and $18 million in today's dollars.

3 Data

3.1 Federal Land Tract Books

We observe the date and precise location of settlement for nearly 200,000 initial homesteaders in Oklahoma from the Federal Land Tract Books. Oklahoma Territory was originally surveyed using the Public Land Survey System (PLSS). The PLSS creates a grid of six mile by six mile townships. Each township contains 36 sections that are each one square mile or 640 acres. Typically, a homesteader would receive a quarter-section of land amounting to 160 acres. Thus, a section would typically contain four homesteaders while a township would contain roughly 144. In the Tract Books we observe each homesteader's name, township, section, and subsection. Figure 1 maps the sections of these initial homesteaders by the year of settlement. Subfigure 1a maps the first year that a homesteader settled in each section which corresponds closely to the timing and location of the series of land runs and land lotteries by which these lands were opened. Subfigure 1b shows the last year a homesteader settled in each section, illustrating the relatively narrow time window within which the land was settled.

3.2 Oil Wells

We link the Tract Books data to three independent sources of oil well data to identify homesteaders with eventual oil discoveries, as well as the dates of discovery. The data on oil wells come from the U.S. Geological Survey (USGS), the Oklahoma Geological Survey (OGS), and a private company called WellDatabase. OGS and WellDatabase both provide locations for individual wells along with their drill dates. USGS is a bit different; it divides the entire state into a grid of one-square-mile cells and reports the number of wells drilled each year in each cell. In the end, each dataset provides us with a count of oil wells in each one-square-mile section.[1]

Unfortunately, the three data sources do not completely agree about where oil wells were located, when they were drilled, and how many there were (see Figure 2). They are most likely to disagree around the edges of oil fields, which may reflect exploratory wells that came up dry and were thus less likely to be recorded. Since observing just one or two wells may be a sign that the wells came up dry, for the USGS and OGS data we require five or more wells in a section before labeling the section as containing oil. The WellDatabase data has approximately 90 percent fewer wells compared to the other two datasets. Since these data are from a private company that sells data on production and current wells, wells in WellDatabase are more likely to be active for longer and less likely to be dry exploratory wells. Thus for WellDatabase, we

[1] We currently link oil to sections; however, linking wells to subsections may produce a more precise measure.

treat the presence of a single well as indicating the presence of oil. Figure 3 maps Oklahoma Territory and colors each section according to how many data sources indicate the presence of oil in that section. For sections with two or three data sources indicating oil, we are reasonably confident that oil was present. For those with only a single data source, we are less confident.

Figure 2 illustrates how all three oil datasets differ somewhat by zooming in on the Cushing oil field. In each subfigure a hand drawn map from 1918 shows the known boundaries of the Cushing oil field in 1918, along with the PLSS sections in the surrounding area. Subfigure 2a overlays purple dots indicating the location of wells based on the OGS oil dataset, and sections with five or more wells are colored in red. Subfigure 2b uses the USGS oil dataset to color sections in red if they contain five or more wells. Due to the nature of the dataset, we do not observe individual well locations, but we do observe the number of wells drilled within a square mile cell each year. Unfortunately these cells don't align perfectly with the PLSS sections, so we connect each cell to the section that contains it's center. Subfigure 2c overlays purple dots indicating the location of wells based on the WellDatabase dataset, and sections with one or more more wells are colored in red. Each dataset traces the countours of the Cushing field reasonably well, although they don't completely agree. Subfigure 2d colors sections in blue if only one data source indicates oil and colors them in red if two or more sources indicate oil. Disagreements tend to arise around the boundaries of the field. Still, Figure 2 illustrates how homesteaders who happened to live inside an oil field could have neighbors just a mile away without oil. And it is this comparison of neighbors which is at the core of our identification strategy.

Within settled Oklahoma Territory the majority of oil was discovered in the late 1910s and 1920s. Figure 4 displays the number of settled sections where oil was discovered for each year. The measurement of oil displayed is our measure where two or more datasets align. We find hardly any wells discovered before 1910 and more than 500 wells discovered after with a large variation of the discovery year. The location of these wells is mapped in figure 5 by decade according to their discovery dates. This figure shows a large variation by location and year of discovery.

3.3 Children of Homesteaders

After observing which homesteaders settled above oil fields, we need to link homesteaders to their children's outcomes. We accomplish this by linking individuals in the Federal Land Tract Books to the 1900 and 1910 U.S. Censuses, following homesteaders across all census years (1850-1940) to compile all of their children, and linking their children forward to the 1940 Census.

Many record linking applications in economics have the advantage of using a broad set of demographic characteristics to establish a unique link. In our setting we only have the names and locations of homesteaders, presenting a challenge in sparse record linking. Fortunately, we also observe the precise locations of their neighbors. After using a supervised machine learning algorithm to assign an ordinal match score to each potential link between the Tract Books and the Census, we rely on information from neighbors to produce reliable links.[2] Figure 6 illustrates this using an example, where Frank Johnson is seen living near several of his neighbors on the same census page although several other Frank Johnson's are observed in the county.

Our linking methods allow us to match 72,603 homesteaders to the 1900 or 1910 U.S. Census, corresponding to a match rate of 39% (see Table 1). Using data from the Bureau of Land Management's General Land Office, however, we find that only 61% of initial claims in Oklahoma were successfully converted to land patents within 5 years. This figure then provides an upper bound on our match rate that is only reached if there is no out-migration after homesteads are purchased.

Table 2 reports summary statistics for the sample of homesteaders linked to the census. Comparing across oil treatment status within the same township we find no statistically significant differences in mean values. Homesteaders were around 36 years old in 1900 and tended to be white, male, married, and farmers.

We identify children of homesteaders by using the Census Tree project, which links individuals across all available census years from 1850 to 1940 (Buckles et al., 2023a). We then link the full set of children to the 1940 U.S. Census. Each homesteader in our sample contributes around three children to the final sample (see Table 2). Table 3 shows the summary statistics of the sample of children. There are no significant differences in children's gender, race, or age across treated and non-treated individuals within the same county. We also report raw differences in outcomes, focusing on measures of wealth (home ownership, home value, and non-wage income), earnings and labor supply (wage income, weekly hours worked, and labor force participation), migration (living in Oklahoma, living in a city), and human capital (years of education).

Our empirical strategy involves comparing children of homesteaders with oil on their land to children of their neighbors who did not have oil. Table 3 shows that 6,242 children in the sample are treated or possibly treated by oil. In townships with oil, 29,737 children of homesteaders are not treated by oil (this statistic is not part of Table 3).

4 Empirical Strategy

Our identifying assumption is that the presence or absence of oil on a homesteader's land is uncorrelated with any unobserved characteristics of the homesteader. Although we cannot directly test this assumption, we do verify that homesteaders with oil are observably similar to their neighbors without oil (see Table 2). Additionally, we find that the presence of oil is unrelated to the likelihood of linking homesteaders to the census. Taken together, this evidence supports our assumption that the presence or absence of oil among early homesteaders was as good as randomly assigned.

[2] In addition to using neighbors, we also utilize the 1890 Oklahoma Territorial Census as a bridge record that contains personally identifiable information for the earliest homesteaders from the 1889 land run.

To estimate the causal effect of an oil discovery, we compare homesteaders with oil discoveries to their neighbors who never experienced an oil discovery. The regression specification is

$$y_{ij} = \delta_1 Oil_j + \delta_2 PossibleOil_j + X_i\beta + a_{T(j)} + \epsilon_{ij}$$

where y_{ij} is an outcome in 1940 for child i of homesteader j, Oil_j is a dummy equal to 1 if two or more data sources indicate the presence of oil, and $PossibleOil_j$ is a dummy variable equal to 1 if only one data source indicates the presence of oil. X_i are covariates[3] and $a_{T(j)}$ is a set of township fixed effects.[4] Intuitively, our estimator compares the children of homesteaders with oil to the children of their neighbors who did not have oil. Because we assign treatment at the section level, standard errors are clustered by sections. Of course, some homesteaders moved away before oil was discovered on their land, so by including all initial homesteaders, we estimate the intention-to-treat effect of oil on child outcomes.

While the current empirical strategy does not require homesteaders to remain on their land until the time of oil in their township, a future version of this paper will use census residence places and land deeds to more closely resemble an estimation of treatment on the treated. Additionally, there may be substantial heterogeneity in the size of the wealth shock as noted in the background section. We plan to use information on oil production amounts and prices to approximate the elasticities of parental wealth.

4.1 Using IV to address measurement error

There are good reasons to worry that our OLS estimates are attenuated due to measurement error in whether or not oil is present. First, measurement error in a binary regressor will bias the OLS coefficient to zero (Lewbel, 2007). In our case, we have three data sources on the location and drill dates of oil wells, all of which disagree somewhat. The measurement error arises for three reasons. First, a given dataset might incorrectly label some homesteader as having no oil by simply missing some oil wells. Second, a dataset might incorrectly label some homesteaders as having oil by recording a well in the wrong location or recording wells which were "dry" and never produced. Third, we might mistakenly attribute oil to a homesteader who lives in a section with oil, but who does not have oil himself.

If the measurement error were classical, then we could obtain a consistent estimate by using two of the noisy measures as instruments for the third. But in the case of a binary regressor, measurement error is non classical so the IV estimator is also inconsistent. Fortunately though, the IV estimator overstates the true effect, so we can use the OLS and IV estimates to bound

[3] As controls, we include gender, race, and age dummies.

[4] The Public Land Survey System (PLSS) divides the land into six mile by six mile "townships", which are further divided into 36 one-square-mile "sections." Typically, each section would contain four homesteaders with 160 acres of land each, so each township contains roughly 144 homesteaders.

the true effect of oil (Kane et al., 1999).[5]

5 Results

5.1 OLS Results

Tables 4 through 6 report estimates of the effects of the presence of oil on the children of homesteaders, relative to children of neighboring homesteaders. These effects are synonymous with the effects of the intention to treat effects of parental wealth, because we do not remove migrants from the sample. Because the gender, race, and age distributions could differ between the treated and control samples (see Table 3), our estimation is a weighted average of within-group effects for the full set of dummies for these variables. Since we analyze outcomes such as wealth and labor supply, we restrict the sample to children who were 18 or older in the 1940 Census (although most of the sample satisfies this requirement).

In addition to estimating effects for the full sample, we add an interaction term for children who were over 18 years old when oil was discovered in their parent's section. This allows us to interpret the main treatment effect as the effect on children who were 18 or younger at the time of treatment, which approach is relevant for understanding childhood effects of wealth.[6] For all regressions we include two classifications of treatment that correspond to the "Oil" and "PossibleOil" variables in our estimating equation, allowing us to compare the main treatment variable to the control group where we are confident that no oil existed by 1940.

Table 4 shows the effects of oil exposure on children's wealth-related outcomes. We find a 2 percentage point increase in the likelihood of home ownership, relative to a control group mean of 52%. This appears to be largely driven by an earlier age of exposure, which the effect increasing to 3 percentage points for children 18 or younger at the time of oil discovery. We also find a 1 percentage point increase in the likelihood of having a home value greater than $5,000, which is only true for 4% of the untreated sample. Nonwage income greater than $50 could contain royalty payments from oil, so this is an interesting (although noisy) indicator.[7] Here we find a 2 percentage point increase in the likelihood of nonwage income relative to a control group mean of 22%. This effect appears to be driven by individuals who were adults when oil was discovered on their parent's land.

We now turn to effects of oil exposure on children's wage income and labor supply which are

[5]As both Kane et al. (1999) and Lewbel (2007) show, we can actually use multiple noisy measures of the presence of oil to consistently estimate the treatment effect using a GMM estimator. We plan to implement this estimator in a future version of the paper.

[6]An earlier age at treatment could be related to two factors as well, which we plan to pursue in more detail. First, it is directly correlated with having an earlier oil discovery date, which could yield more wealth due to higher oil prices prior to 1930. Second, it reduces the likelihood of a child of a homesteader also being a homesteader to zero because oil discoveries occurred after homestead claims (which were only possible for adults).

[7]The indicator for nonwage income is set to zero for farmers, who may have counted their earnings as nonwage income.

reported in Table 5. We find no effects on wage income or labor force participation for the full sample. However, we do find that children who were younger at the time of oil discovery worked 2.12 fewer hours per week than children without oil on their parent's land.

Our final set of outcomes are related to education and migration. We find no effects on educational attainment or the likelihood of children living in (or leaving) Oklahoma as adults. However, oil exposure does lead to a 3 percentage point increase in the likelihood of living in a city relative to a control group mean of 38%, which we interpret as a proxy for urban migration. While it is possible that oil discoveries could have caused urban areas to spring up around homesteaders with oil, we consider this unlikely. Recall that we are comparing these homesteaders to their close neighbors, so we would expect any urban growth around the oil discovery to also affect those neighbors. Moreover, this effect is largely concentrated on children who were 18 or younger at the time of oil discovery, which is consistent with the idea of parental wealth allowing children to migrate to urban areas.

5.2 IV Results

In Tables 7–9, we report OLS and IV estimates using the OGS measure of oil as our independent variable and the USGS and WellDatabase measures as instruments. For each dependent variable, we report the OLS coefficient(s) in the left column and the IV coefficient(s) in the right column. We also report estimates from a specification that interacts the presence of oil with a dummy for whether the child was over 18 years old at the time of oil discovery. In general, we find similar effect sizes to those reported above. The bounds provided by the OLS and IV estimates tend to be relatively narrow and typically contain the estimates from Tables 4–6. One major exception is years of education. The OLS estimate (using the OGS data only) finds that oil raises years of education by 0.4 years for children who were under 18 at the time of discovery. The IV estimate is nearly identical, although the standard error rises by enough to make it statistically insignificant. This is in contrast with our earlier result that having 2 or more datasets indicating oil raised education by (a statistically insignificant) 0.18 years. We also find somewhat stronger effects of oil on migrating to an urban area and nonwage income over $50.

6 Conclusion

We leverage the quasi-random discovery of oil discoveries in early twentieth century Oklahoma to estimate the long run effects of parental wealth shocks on child outcomes. Oklahoma was settled quickly and homesteaders were unaware of the location or existence of oil when they settled. We compare several long run outcomes of children of homesteaders who were fortunate enough to have oil with the children of their neighbors who were not so lucky. We find that having oil raised the probability of children owning a home, having non-labor income, and living in a city. Children with oil were more likely to live in a city in 1940, especially those who were younger when oil was discovered. Children who were younger at the oil discovery also work 2 fewer hours per week. Notably, we do not find an effect on earnings or education

Our results inform a large literature that has focused on estimating the correlation between parent income and wealth on child outcomes. We build on a much smaller literature that has used lotteries to estimate the causal effects of parent wealth on children's outcomes (Bulman et al., 2021; Bleakley and Ferrie, 2016). Relative to these papers, our data allow us to use a larger dataset and look at a wider range of long run outcomes including education, income, and measures of wealth.

In addition, our paper builds on the Census Tree project (Buckles et al., 2023a) by linking original Oklahoma homesteaders to the Census Tree. Despite having no demographic information on homesteaders, we still achieve a high match rate by comparing the names of neighbors in the Tract Books and neighbors on the census sheets. Our linking method can be applied to other "sparse linking" settings and illustrates the potential for using detailed geographic information to obtain high quality matches.

References

Philipp Ager, Leah Boustan, and Katherine Eriksson. The intergenerational effects of a large wealth shock: White southerners after the civil war. *American Economic Review*, 111:3767–3794, 11 2021.

Hoyt Bleakley and Joseph Ferrie. Shocking behavior: Random wealth in antebellum georgia and human capital across generations. *The Quarterly Journal of Economics*, pages 1455–1495, 2016.

Kasey Buckles, Adrian Haws, Joseph Price, and Haley E.B. Wilbert. Breakthroughs in Historical Record Linking Using Genealogical Data: The Census Tree Project. *NBER Working Paper Series*, 2023a.

Kasey Buckles, Joseph Price, Zachary Ward, and Haley Wilbert. Family trees and falling apples: Intergenerational mobility estimates from u.s. genealogy data. *Working paper*, 2023b.

George Bulman, Robert Fairlie, Sarena Goodman, and Adam Isen. Parental resources and college attendance: Evidence from lottery wins. *American Economic Review*, 11:1201–1240, 4 2021.

Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129:1553–1623, 2014.

Dylan Shane Connor and Michael Storper. The changing geography of social mobility in the united states. *PNAS*, 117, 2020.

James J. Feigenbaum. Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. *The Economic Journal*, 128:F446–F481, 2018.

Thomas J Kane, Cecilia Elena Rouse, and Douglas Staiger. Estimating returns to schooling when schooling is misreported. Working Paper 7235, National Bureau of Economic Research, July 1999. URL `http://www.nber.org/papers/w7235`.

Arthur Lewbel. Estimation of Average Treatment Effects with Misclassification. *Econometrica*, 75(2):537—-551, 2007.

Claudia Olivetti and M. Daniele Paserman. In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940. *American Economic Review*, 105: 2695–2724, 2015.

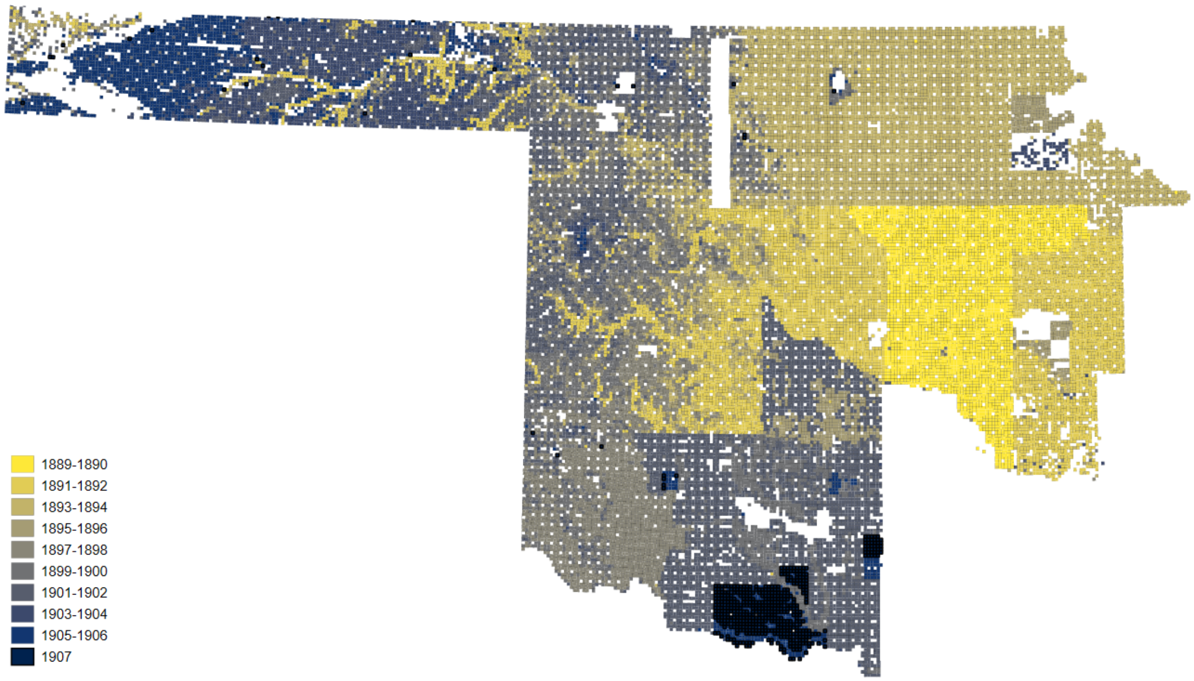
Xi Song, Catherine G Massey, Karen A Rolf, Joseph P Ferrie, Jonathan L Rothbaum, and Yu Xie. Long-term Decline in Intergenerational Mobility in the United States Since the 1850s. *Proceedings of the National Academy of Sciences*, 111:251–258, 10 2019.

Zachary Ward. Intergenerational Mobility in American History: Accounting for Race and Measurement Error. *American Economic Review (forthcoming)*, 2023.

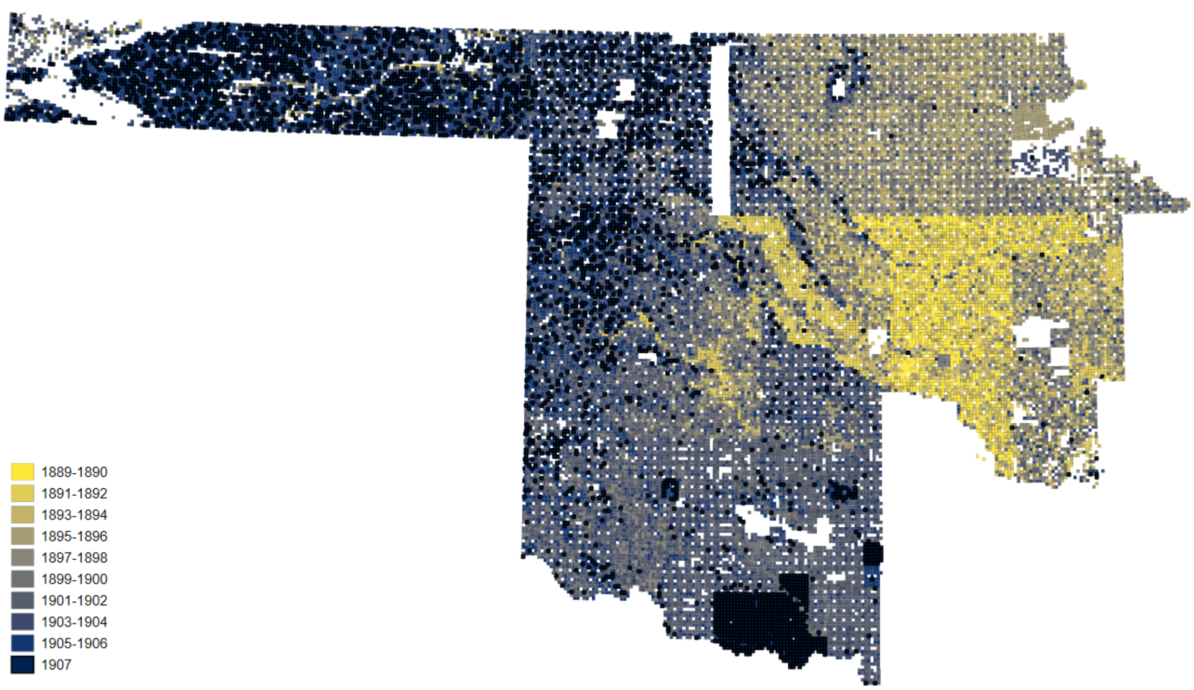
7 Tables and Figures

Figure 1: Oklahoma Territory - Homesteading Years

(a) First year of homesteading within a section



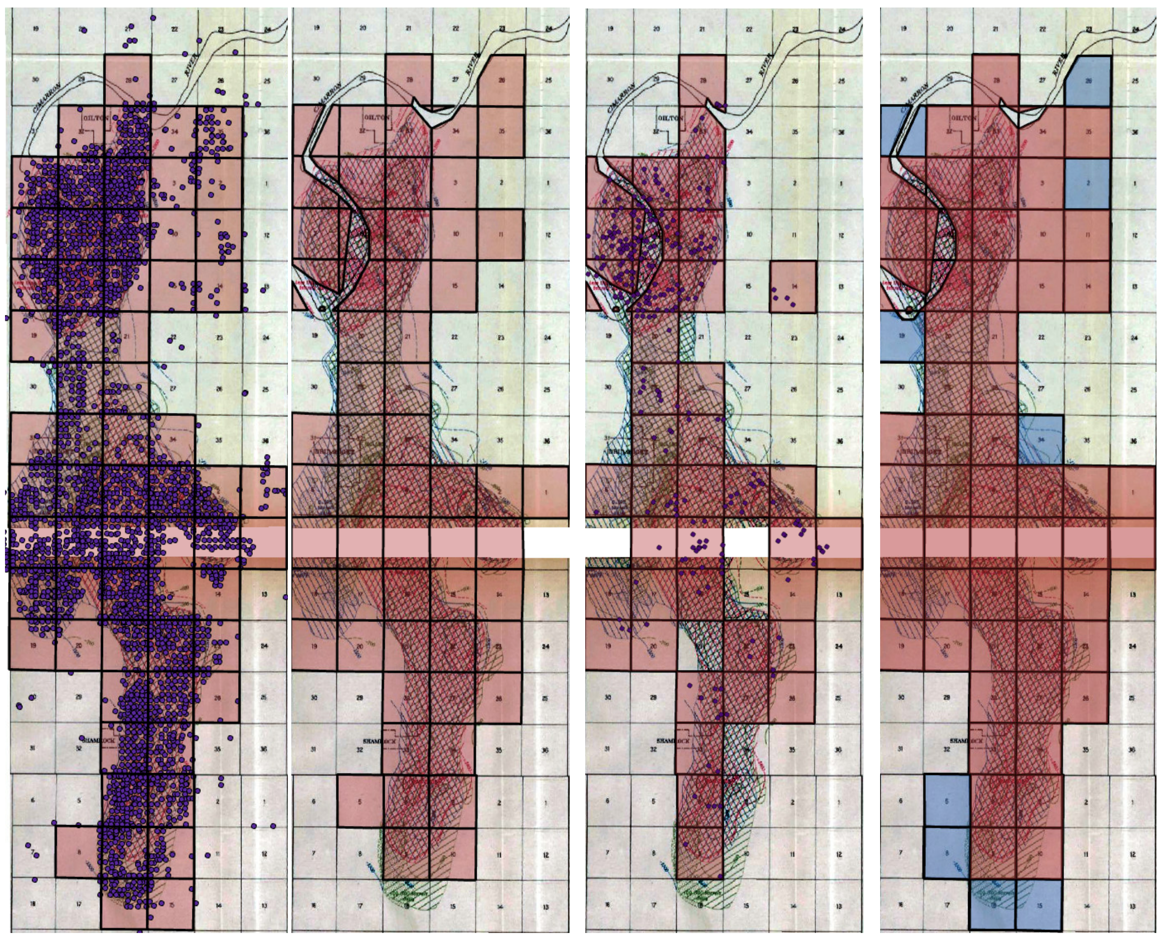
(b) Last year of homesteading within a section



Notes: Panel (a) maps out the first year a 1889-1907 homesteader settled within their section in Oklahoma Territory. Panel (b) maps out the last year a 1889-1907 homesteader settled within their section in Oklahoma Territory. Year of homesteading and section is given by the Federal Land Tract Books.

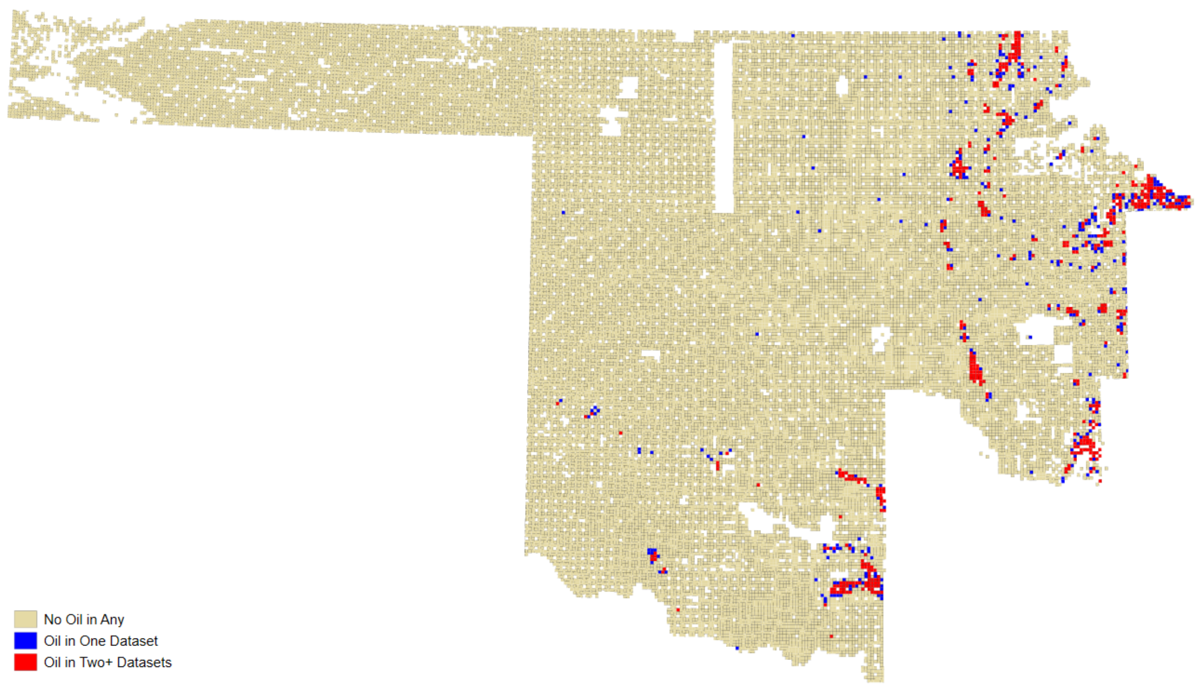
Figure 2: Cushing Oil Field Map by 1918

(a) Oklahoma Geological Survey (b) U.S. Geological Survey (c) WellDatabase (d) All Oil



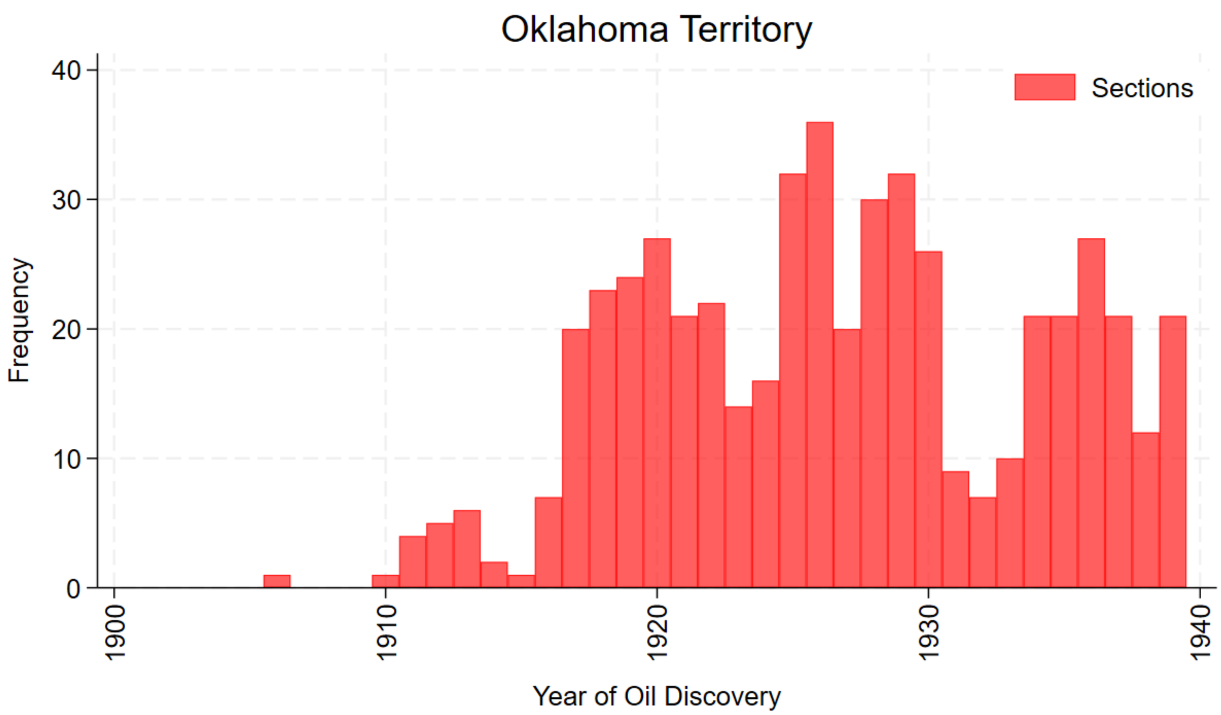
Notes: Historical sketch map of Cushing Oil Field is by Carl H. Beal through the United States Geological Survey in 1918. Red squares in (a) represent five or more wells drilled within the section before 1918. Red squares in (b) represent five or more wells drilled before 1918 within the one square mile area. Red squares in (c) represent one or more wells drilled before 1918. Purple dots map out oil wells in (a) and (c). Blue squares in (d) represent sections where oil is present in only one dataset and red squares represent sections where oil is present in two or three datasets.

Figure 3: Sections in Oklahoma Territory with Oil



Notes: Tan areas represent sections where homesteaders settled between 1889-1907 and no oil is present before 1940. Blue areas represent sections where homesteaders settled between 1889-1907 and oil is present in only one dataset before 1940. Red areas represent sections where homesteaders settled between 1889-1907 and oil is present in two or three datasets before 1940. White areas represent sections where no homesteaders initially settled due to the land being reserved for school funds or for Indian settlements.

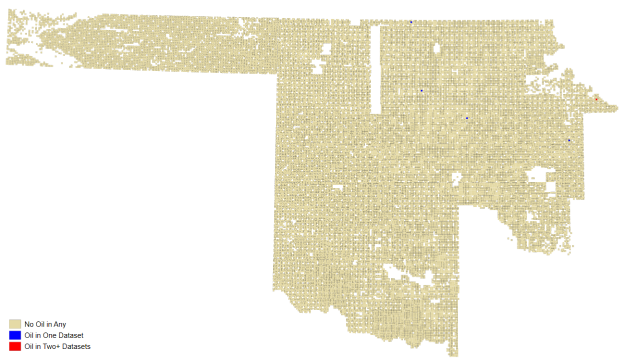
Figure 4: Year of Oil Discovery in Homesteaded Oklahoma Territory



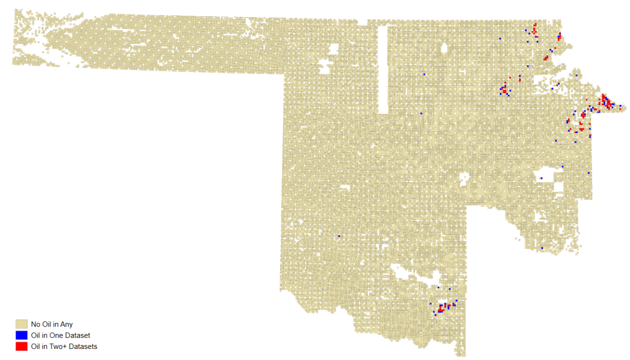
Notes: This figure plots the number of sections where oil was discovered in each year within settled Oklahoma Territory. We define an oil discovery to occur when two or more data sources indicate the presence of oil in the section.

Figure 5: Oil Discoveries In Oklahoma Territory by Decade

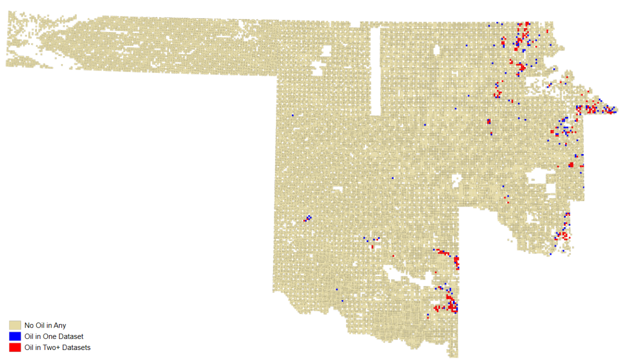
(a) Oil Before 1907



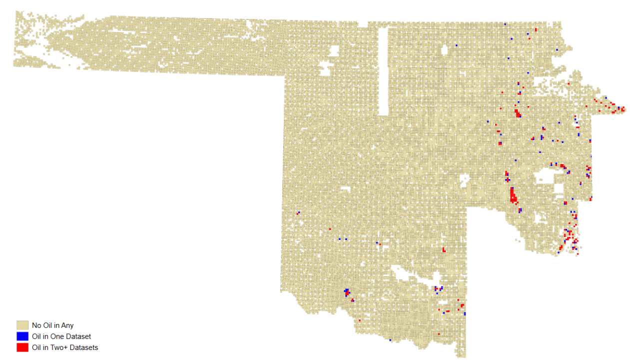
(b) Oil Between 1907-1919



(c) Oil Between 1920-1929



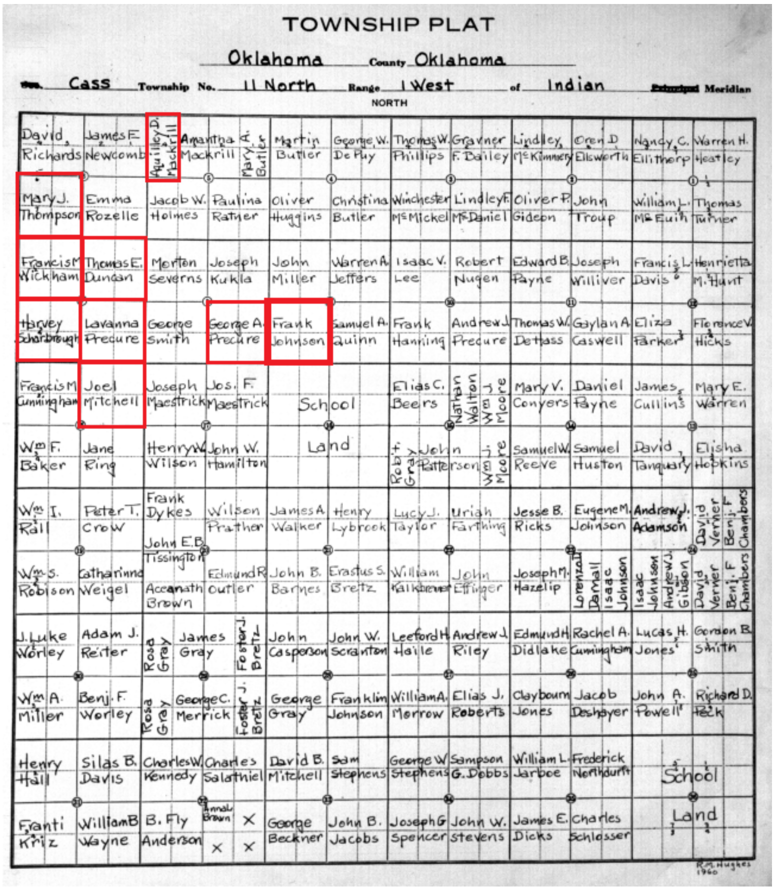
(d) Oil Between 1930-1939



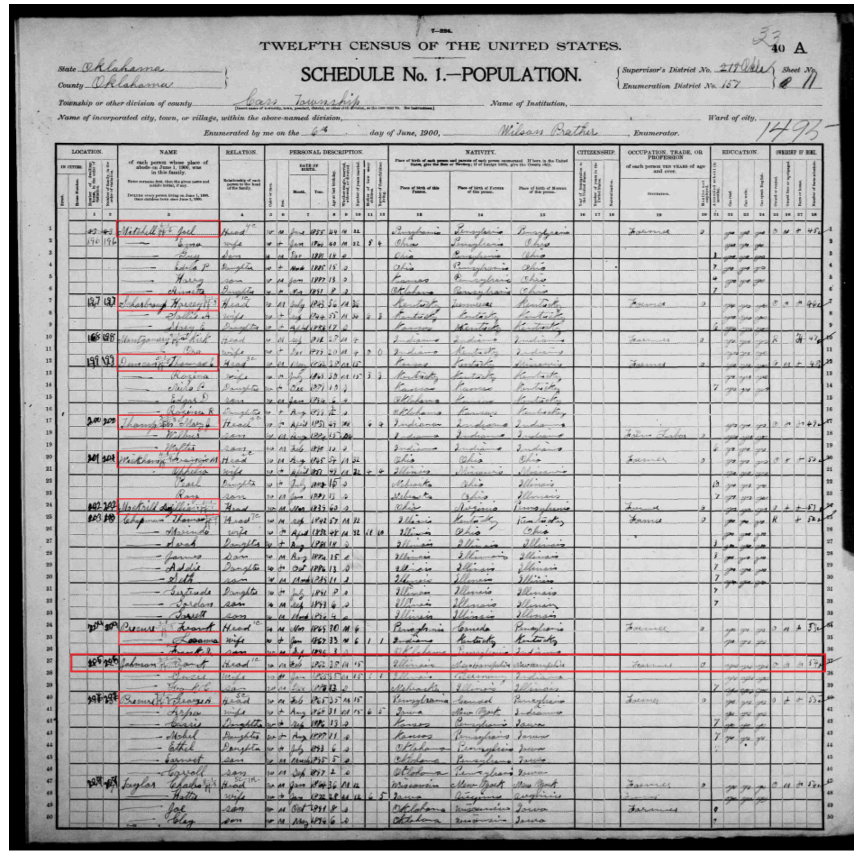
Notes: The year of oil discovery in two plus datasets is measured by the second lowest year of oil discovery. Tan areas represent sections where homesteaders settled between 1889-1907 and no oil is present in the years specified. Blue areas represent sections where homesteaders settled between 1889-1907 and oil is present in only one dataset within the years specified. Red areas represent sections where homesteaders settled between 1889-1907 and oil is present in two or three datasets in the years specified. White areas represent sections where no homesteaders initially settled due to the land being reserved for school funds or for Indian settlements.

Figure 6: Example of linking homesteaders to the U.S. Census

(a) Plat map



(b) Census page



Notes: Panel (a) shows the location of Frank Johnson and several neighbors in a township plat map in Oklahoma County, available from okhistory.org. Panel (b) shows where these neighbors can be co-located on a page of the 1900 U.S. Census in Oklahoma County, downloaded from FamilySearch.org.

Table 1. Summary statistics by linkage to 1900 or 1910 census, homesteaders

	Linked	Unlinked	Within township difference
Intention to treat			
Oil in section (2-3 sources)	0.02	0.02	-0.02
	[0.13]	[0.13]	(0.01)
Oil in section (1 source only)	0.01	0.01	0.00
	[0.10]	[0.10]	(0.01)
Tract books			
Year of entry	1897.45	1898.41	0.03**
	[5.07]	[5.26]	(0.01)
Male first name	0.84	0.78	0.04***
			(0.00)
Given names count	446.39	430.11	-0.87
	[735.54]	[721.93]	(3.57)
Surname count	115.53	186.97	-73.64***
	[284.91]	[381.05]	(1.63)
Has middle initial	0.69	0.70	-0.02***
			(0.00)
Observations	72,603	113,290	

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Means reported for initial homesteaders in Oklahoma Territory, with standard deviations in brackets. Within township differences are computed as coefficients from a series of regressions with an intention to treat variable as a dependent variable and a dummy for being linked as an independent variable, with township fixed effects and standard errors (in parentheses) clustered by section. A township comprises 36 sections, each with one square mile. "Given names count" is a measure of name commonality, computed as the number of homesteaders who share the same given names. The same is true for "surname count".

Table 2. Summary statistics, linked homesteaders

	Oil in section (2-3 sources)	Oil in section (1 source only)	No oil	Difference (I - III)		Difference (II - III)	
	(I)	(II)	(III)	Within township	Unconditional	Within township	Unconditional
Year of entry	1894.93	1894.51	1897.53	0.03	-2.60***	0.06	-3.02***
	[3.89]	[3.76]	[5.08]	(0.08)	(0.11)	(0.10)	(0.14)
Male	0.89	0.86	0.85	0.01	0.04***	-0.02	0.01
				(0.01)	(0.01)	(0.01)	(0.01)
Black	0.04	0.03	0.04	-0.01	0.00	-0.02*	-(0.01)
				(0.01)	(0.01)	(0.01)	(0.01)
Indigenous	0.01	0.02	0.02	0.00	0.00	0.01	0.01
				(0.00)	(0.00)	(0.01)	(0.01)
Age	36.27	35.60	32.47	-0.27	3.81***	-0.97	3.13***
	[17.47]	[18.20]	[18.35]	(0.59)	(0.48)	(0.71)	(0.65)
Married	0.72	0.70	0.67	0.00	0.04***	-0.03	0.03
				(0.02)	(0.01)	(0.02)	(0.02)
Home	0.77	0.76	0.76	0.01	0.01	0.00	0.00
				(0.01)	(0.01)	(0.02)	(0.02)
On farm	0.73	0.77	0.78	-0.02	-0.05***	0.01	-0.01
				(0.01)	(0.01)	(0.02)	(0.02)
Farmer	0.68	0.68	0.63	0.01	0.05***	0.00	0.05***
				(0.02)	(0.01)	(0.02)	(0.02)
Children linked to 1940	3.12	3.11	3.22	0.08	-0.10	0.02	-0.11
	[2.78]	[2.70]	[2.83]	(0.09)	(0.08)	(0.11)	(0.10)
Observations	1,336	791	70,476				

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Means reported for initial homesteaders in Oklahoma Territory, with standard deviations in brackets. We report unconditional differences in means with standard errors in parentheses, as well as coefficients from a series of regressions with each characteristic as a dependent variable, one of the treatment variables and township fixed effects as independent variables, and standard errors clustered by section (comparing to sections with no oil). Socioeconomic characteristics are obtained from linkages to the 1900 and 1910 census (whichever is earliest). Section-level oil is assigned using drilled wells from three sources for which each has a different cutoff (as described in the paper). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 3. Summary statistics in 1940 census, children of homesteaders

	Oil in section (2-3 sources)	Oil in section (1 source only)	No oil	Difference (I - III)		Difference (II - III)	
	(I)	(II)	(III)	Within township	Unconditional	Within township	Unconditional
Characteristics							
Male	0.52	0.52	0.51	0.00	0.01	0.00	0.01
				(0.01)	(0.01)	(0.01)	(0.01)
Black	0.02	0.02	0.02	0.00	0.00	-0.01	0.00*
				(0.00)	(0.00)	(0.01)	(0.00)
Indigenous	0.01	0.01	0.01	0.00	0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)	(0.00)
Age	45.39	45.72	42.73	-0.09	2.66***	0.37	3.00***
	[16.83]	[17.58]	[17.57]	(0.48)	(0.27)	(0.64)	(0.37)
Over 18 years at oil discovery	0.68	0.73					
Outcomes							
Home	0.55	0.54	0.52		0.04***		0.02*
					(0.01)		(0.01)
Home value >$5000	0.06	0.06	0.04		0.02***		0.01***
					(0.00)		(0.00)
Nonwage income >$50	0.24	0.23	0.21		0.04***		0.03***
					(0.01)		(0.01)
Wage income	389.83	358.32	322.69		67.14***		16.25**
	[763.69]	[726.42]	[663.70]		(13.09)		(16.25)
Hours worked	18.15	18.44	18.98		-0.83**		-0.54
	[24.64]	[25.01]	[24.98]		(0.04)		(0.52)
In labor force	0.52	0.52	0.54		-0.02**		-0.02
					(0.01)		(0.01)
Oklahoma	0.54	0.53	0.49		0.05***		0.03***
					(0.01)		(0.01)
Urban	0.42	0.41	0.36		0.06***		0.04***
					(0.01)		(0.01)
Years of Education	8.65	8.62	8.53		0.13**		0.09
	[3.46]	[3.36]	[3.42]		(0.06)		(0.07)
Observations	3,932	2,310	212,329				

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Means reported for children of initial homesteaders in Oklahoma Territory, with standard deviations in brackets. We report unconditional differences in means with standard errors in parentheses, as well as coefficients from a series of regressions with each characteristic as a dependent variable, one of the treatment variables and township fixed effects as independent variables, and standard errors clustered by section (comparing to sections with no oil). Variables are obtained from linkages to the 1940 census. Section-level oil is assigned using drilled wells from three sources for which each has a different cutoff (as described in the paper). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 4. Effects of oil exposure on children's wealth in 1940

	Home		Home value >$5000		Nonwage income >$50	
Oil in section (2-3 sources)	0.02**	0.03*	0.01**	0.01	0.02**	-0.01
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Oil in section (2-3 sources) x over 18 at discovery		-0.02		0.00		0.03**
		(0.02)		(0.01)		(0.02)
Oil in section (1 source only)	0.01	0.00	0.01	0.01	0.00	0.00
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Oil in section (1 source only) x over 18 at discovery		0.02		-0.01		0.00
		(0.02)		(0.01)		(0.02)
Mean outcome, no oil	0.52		0.04		0.22	
Observations	198,690		198,690		198,690	

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: OLS models with township fixed effects and standard errors (in parentheses) clustered by section. Fixed effects are included for gender, race, and age. Sample includes individuals age 18 or older in 1940 whose parent was an initial homesteader in Oklahoma Territory. Treatment is defined using drilled wells from three sources for which each has a different cutoff (as described in the paper). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 5. Effects of oil exposure on children's income and labor supply in 1940

	Wage income		Hours worked		In labor force	
Oil in section (2-3 sources)	-0.88	15.55	-0.59	-2.12***	-0.01	-0.02
	(17.45)	(27.29)	(0.44)	(0.80)	(0.01)	(0.01)
Oil in section (2-3 sources) x over 18 at discovery		-22.80		1.96**		0.02
		(29.93)		(0.84)		(0.01)
Oil in section (1 source only)	-26.92	-28.03	0.15	-0.68	0.00	0.00
	(19.93)	(31.05)	(0.50)	(0.97)	(0.01)	(0.01)
Oil in section (1 source only) x over 18 at discovery		1.81		1.05		-0.01
		(30.30)		(1.00)		(0.01)
Mean outcome, no oil	353.23		20.52		0.53	
Observations	179,122		198,690		198,690	

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: OLS models with township fixed effects and standard errors (in parentheses) clustered by section. Fixed effects are included for gender, race, and age. Sample includes individuals age 18 or older in 1940 whose parent was an initial homesteader in Oklahoma Territory. Treatment is defined using drilled wells from three sources for which each has a different cutoff (as described in the paper). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 6. Effects of oil exposure on children's education and migration in 1940

	Years of Education		Oklahoma		Urban	
Oil in section (2-3 sources)	0.02	0.18	0.01	-0.02	0.03**	0.06***
	(0.10)	(0.16)	(0.02)	(0.03)	(0.01)	(0.02)
Oil in section (2-3 sources) x over 18 at discovery		-0.21		0.04*		-0.04*
		(0.16)		(0.02)		(0.02)
Oil in section (1 source only)	0.03	-0.01	0.00	-0.04	0.01	0.03
	(0.10)	(0.16)	(0.02)	(0.03)	(0.01)	(0.03)
Oil in section (1 source only) x over 18 at discovery		0.05		0.05**		-0.03
		(0.15)		(0.03)		(0.02)
Mean outcome, no oil	8.84		0.49		0.38	
Observations	192,753		198,690		198,690	

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: OLS models with township fixed effects and standard errors (in parentheses) clustered by section. Fixed effects are included for gender, race, and age. Sample includes individuals age 18 or older in 1940 whose parent was an initial homesteader in Oklahoma Territory. Treatment is defined using drilled wells from three sources for which each has a different cutoff (as described in the paper). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 7. Effects of oil exposure on children's wealth in 1940 using instrumental variables

	Home		Home Value > $5000		Nonwage Income > $50	
	(1)	(2)	(1)	(2)	(1)	(2)
OGS oil in section	0.010	0.052***	0.010*	0.030***	0.014	0.033***
	(0.011)	(0.016)	(0.006)	(0.010)	(0.009)	(0.013)
OGS oil in section	0.009	0.045	0.010	0.032	0.010	-0.012
	(0.021)	(0.032)	(0.010)	(0.020)	(0.015)	(0.025)
OGS oil in section × over 18 at discovery	0.001	0.008	0.001	-0.002	0.005	0.056**
	(0.021)	(0.032)	(0.009)	(0.018)	(0.016)	(0.028)
IV	-	X	-	X	-	X
Mean, no Oil	0.52		0.04		0.22	
Observations	198,678		198,678		198,679	

*$p < .1$, **$p < .05$, ***$p < .01$

Notes: OLS and two-stage least-squares models with township fixed effects and standard errors (in parentheses) clustered by section. Instrumental variables include Well Database oil in section and USGS oil in section with their respective cutoffs of drilled wells. With regressions separating by age of discovery, instrumental variables include the interaction of oil in section with age 18 or over of child at discovery. Control variables include an indicator for white, an indicator for Native American, age, age squared, and an indicator for gender. Sample includes individuals age 18 or older in 1940 whose parent was an initial homesteader in Oklahoma Territory. Treatment is defined using drilled wells from OGS with its respective cutoff.

Table 8. Effects of oil exposure on children's income and labor supply in 1940 using instrumental variables

	Wage Income		Hours Worked		In Labor Force	
	(1)	(2)	(1)	(2)	(1)	(2)
OGS oil in section	0.275	-13.080	-0.470	-0.316	-0.008	-0.008
	(16.300)	(23.195)	(0.411)	(0.592)	(0.007)	(0.009)
OGS oil in section	25.055	61.213	-0.436	-2.735**	-0.003	-0.026
	(27.411)	(47.384)	(0.813)	(1.344)	(0.014)	(0.020)
OGS oil in section × over 18 at discovery	-30.474	-92.715*	-0.041	2.986**	-0.007	0.022
	(29.130)	(50.013)	(0.848)	(1.424)	(0.014)	(0.022)
IV	-	X	-	X	-	X
Mean, no Oil	353.23		20.52		0.53	
Observations	179,112		198,679		198,679	

*$p < .1$, **$p < .05$, ***$p < .01$

Notes: OLS and two-stage least-squares models with township fixed effects and standard errors (in parentheses) clustered by section. Instrumental variables include Well Database oil in section and USGS oil in section with their respective cutoffs of drilled wells. With regressions separating by age of discovery, instrumental variables include the interaction of oil in section with age 18 or over of child at discovery. Control variables include an indicator for white, an indicator for Native American, age, age squared, and an indicator for gender. Sample includes individuals age 18 or older in 1940 whose parent was an initial homesteader in Oklahoma Territory. Treatment is defined using drilled wells from the OGS with its respective cutoff.

Table 9. Effects of oil exposure on children's education and migration in 1940 using instrumental variables

	Years of Education		Oklahoma		Urban	
	(1)	(2)	(1)	(2)	(1)	(2)
OGS oil in section	0.077	0.046	0.011	0.013	0.032***	0.049***
	(0.088)	(0.127)	(0.014)	(0.021)	(0.012)	(0.018)
OGS oil in section	0.404**	0.405	-0.041	-0.027	0.082***	0.089**
	(0.163)	(0.253)	(0.027)	(0.040)	(0.024)	(0.036)
OGS oil in section × over 18 at discovery	-0.398**	-0.448*	0.064**	0.050	-0.061***	-0.050
	(0.158)	(0.254)	(0.025)	(0.039)	(0.023)	(0.036)
IV	-	X	-	X	-	X
Mean, no Oil	8.84		0.49		0.38	
Observations	192,744		198,679		198,679	

*$p < .1$, **$p < .05$, ***$p < .01$

Notes: OLS and two-stage least-squares models with township fixed effects and standard errors (in parentheses) clustered by section. Instrumental variables include Well Database oil in section and USGS oil in section with their respective cutoffs of drilled wells. With regressions separating by age of discovery, instrumental variables include the interaction of oil in section with age 18 or over of child at discovery. Control variables include an indicator for white, an indicator for Native American, age, age squared, and an indicator for gender. Sample includes individuals age 18 or older in 1940 whose parent was an initial homesteader in Oklahoma Territory. Treatment is defined using drilled wells from the OGS with its respective cutoff.