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

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

March 21, 2025

Abstract

What are the causal effects of parental wealth on children? 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 later oil discoveries and develop new methods to link them to their children in the 1940 census, allowing us to examine the long-run impacts of parental wealth shocks on children's wealth, income, labor supply, education, migration, and marital outcomes. We find that oil discoveries increase several proxies for children's wealth, but only among children whose parents are most likely to be deceased by 1940, suggesting that oil wealth is transferred primarily through bequests, rather than inter vivos transfers. Although we fail to find an effect on educational attainment or labor market outcomes for the full sample, we do find that oil discoveries caused children in their late teens and early twenties to enroll in school at higher rates. We also find evidence that parental wealth delayed marriage for daughters and increased the educational attainment of their future husbands.

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

Economists and policymakers have long been interested in understanding the degree to which a person's income or wealth depends on the income or wealth of their parents. 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. (2023c), Ward (2023)) and regions (Chetty et al. (2014), Connor and Storper (2020)). Variation in mobility rates across time and space suggests an important role for national and local policies to shape intergenerational mobility. However, nearly all of these papers estimate correlations instead of causal effects. Identifying the causal effects of parental wealth is challenging because children can be impacted by unobserved parental traits that are correlated with wealth, including health, education, social capital, and cultural values.

We leverage quasi-random oil discoveries in early twentieth-century Oklahoma to estimate the causal long-run impacts of a large wealth shock on children. The first major oil discovery occurred in 1897 near Bartlesville. 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 financial windfalls to those who owned land with oil. We exploit the quasi-random nature of these discoveries by comparing the children of homesteaders fortunate enough to experience oil discoveries with the children of their neighbors who were not as fortunate. Consider the story of Frank Wheeler as recounted in Franks (1980, pp. 68–71).

> By the time Oklahoma entered the Union in 1907, Frank Wheeler had already purchased a farm about twelve miles east of the community of Cushing... There, he shared a small log house with his wife, eight daughters, and one son. The rough terrain was described as "sixty acres of plow land and one hundred acres of rocks and scrub." It was difficult for Wheeler to wrest a living from the barren earth, and he spent much of the time seeking employment as a mason to supplement his income. Even so, the family remained constantly on the brink of poverty....
>
> Having heard of the oil strikes in Oklahoma, Pennsylvanian Tom Slick arrived in Cushing during the winter of 1911.... Slick was known for drilling wildcat wells, and many potential investors were skeptical of his idea of oil in the area, pointing out that the nearest producing well was more than twenty-six miles away....
>
> Although somewhat hesitant, Wheeler, after being informed of Slick's reason for being in the area, agreed to lease the land.

After drilling a well on the Wheeler farm,

> ... on March 12, 1912, it blew in "with a roar of gas-driven oil that literally flooded the surrounding earth." The tremendous pressure of the natural gas in the well forced

> the crude to the surface at a rate of 400 barrels a day. The great Cushing Pool had been discovered. Within a month, Wheeler was collecting $125 daily [$3,267 in 2019 dollars] in royalties from the barren 160 acres that had produced such a meager existence for his family only days previously.

Others in the area also benefited enormously from Slick's discovery.

> Leasing activity in the Cushing area continued unabated. Money flowed in a seemingly unending stream to those fortunate enough to own land near the discovery well. Aaron Drumwright, the owner of 120 acres located south of the Wheeler farm, received $20,000 for property he had purchased for $1,500 only six months previously. R. A. Fulkerson had chopped wood at $2 per cord prior to Slick's find but, after selling his small acreage just south of Drumwright, was able to establish himself as a well-to-do farmer in his native Kentucky. Sara Rector, the orphaned daughter of a Creek slave, inherited her father's 160 acres of land at the age of ten and saw a well drilled on the barren land flow at the rate of three thousand barrels of oil daily. On April 12, 1912, the Prairie Oil and Gas Company paid $64,000 for 40 acres south of Slick's well, even though the land generally sold for approximately $55 an acre.

But not everyone was so fortunate. Depending on where the oil field boundaries happened to fall, some homesteaders struck it rich while others did not. Our empirical strategy compares children of homesteaders like Frank Wheeler, who struck it rich with oil, to the children of their less fortunate neighbors.

We combine data on the locations and drill dates of all oil wells drilled in Oklahoma with land patent records to identify homesteaders who experienced an oil discovery on their land before 1940. We then link each homesteader to the census and use cross-census links from the Census Tree project (Buckles et al., 2023b) to follow children into adulthood. Linking land patents to the census presents a challenge because these records only list the homesteader's name and the location of his or her homestead. We develop a new linking method that exploits the detailed geographic information in the homesteader records, especially information about neighbors, to achieve a high match rate relative to the prior literature. Our method may be useful in other contexts where researchers need to link records to the census that contain sparse demographic data but rich geographic information. By linking land patents to the Census Tree, we are able to follow homesteaders and their families over time and observe the outcomes of children in 1940, which was years or even decades after the oil discovery occurred.

Our empirical strategy identifies the causal effects of parental wealth shocks on children, but the interpretation of our estimates is nuanced. We illustrate this using two thought experiments. First, imagine a wealth shock to a single family and comparing the children's outcomes to the neighbors' children. Second, imagine a wealth shock to all of the family's neighbors, but not the family itself, and comparing their outcomes with children from a similar community where

no one received a wealth shock. The first thought experiment represents a pure "individual effect"—the family receives a wealth shock, but the wealth of their community is unchanged. The second represents a pure "community effect"—the family's wealth is unchanged, but the community around them becomes wealthier. The discovery of oil in Oklahoma falls between these two thought experiments. Some families in a community discovered oil while others did not (an individual effect), and the community as a whole became wealthier, potentially affecting all children in the community, not just those who actually discovered oil (a community effect). For instance, children whose parents did not discover oil may still have benefited from additional funding for the local public school or from local economic growth more generally. By comparing children of homesteaders with oil to their neighbors without oil, we net out any community effect and identify just the individual effect of wealth.

After verifying that oil discoveries are unrelated to observed characteristics of homesteaders, we estimate the causal effect of oil discoveries by comparing the children of homesteaders who were fortunate enough to experience an oil discovery with their neighbors who were not. We find that oil discoveries increase several proxies for children's wealth, but only among children whose parents are most likely to be deceased by 1940. This suggests that oil wealth is transferred primarily through bequests from parents to children, rather than through *inter vivos* transfers. We also find evidence that wealth shocks affect outcomes beyond children's monetary wealth. Oil discoveries caused children between ages 14–24 to enroll in school at higher rates in 1940. Although we find no effect on labor market participation, wages, or hours worked, we do find suggestive evidence that children who were young at the time of discovery are more likely to work in white collar occupations and that sons are more likely to live in a city. We find that oil discoveries cause daughters to delay marriage and to ultimately marry a more educated spouse, suggesting that parental wealth gives daughters greater freedom to delay marriage in search of a better match.

As the story of Frank Wheeler illustrates, oil discoveries could be quite large, though they did vary substantially in size. 6.8 percent of the children we identify as having an oil discovery (based on wells drilled) do not have any oil production recorded.[1] Among those with positive production, median total production is 76,335 barrels which is equivalent to royalties of $247,524 in 2023 dollars. Production varies substantially with a long right tail—mean production is 475,484 barrels equivalent to royalties of $1.4 million. Focusing on those with above-median production volumes, production averages 933,152 barrels ($2.7 million in royalties). For comparison, annual earnings average just $16,880 among children with no oil discoveries. In other words, these oil discoveries represented a potentially life-changing amount of money for the families in our sample.

A small number of papers have also used natural experiments to estimate the causal effects

[1] As we explain in the paper, for data quality reasons we prefer to use well counts to assign treatment rather than production volumes.

of parental wealth shocks on long-run 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 attained 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 larger sample, including a sample of daughters in addition to sons. 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, the Georgia Land Lottery and emancipation occurred in the context of a nineteenth century agrarian society but Oklahoma oil discoveries occurred within the context of a twentieth-century industrialized economy. Fourth, oil discoveries 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.

Another group of papers has examined the effect of oil and gas discoveries on landowners. For example, Cookson et al. (2022) estimate how windfalls from fracking influence household debt repayment, while Bellon et al. (2021) explore their impact on business formation. Our empirical strategy also compares landowners with oil discoveries to neighboring landowners without such discoveries. However, rather than focusing on the immediate effects on the landowners, we investigate the long-run impacts on their children. Moreover, our study is set in the early twentieth century and examines wealth shocks that are an order of magnitude larger.[2]

Also related to our study is a literature on the local economic effects of discovering and extracting natural resources. Black et al. (2005b) find the Appalachian coal boom reduced high school enrollment rates by raising the opportunity costs of schooling. In a review of the literature, Mousavi and Clark (2021) report that a majority of studies find negative effects of local resource extraction on human capital investment. Looking at local economic effects more broadly, Black et al. (2005a) find employment spillovers from the Appalachian coal boom to sectors with locally traded goods, though not to nationally traded goods. Feyrer et al. (2017) also find substantial increases in local wage income due to fracking. Muehlenbachs et al. (2015) report that shale discoveries affect house prices differently depending on whether the homes relied on groundwater or not. Cust and Poelhekke (2015) review the literature and find evidence for positive direct economic effects accompanied by negative effects on governance and corruption. In contrast with these papers, we do not focus on effects at the community-level. Rather, by comparing

[2]While the largest wealth shocks reported by Cookson et al. (2022) amount to $46,245 (at the 99th percentile), our median wealth shock is $247,524 with a mean of $1.4 million (in 2023 dollars).

close neighbors in the same community, we effectively net out community effects and focus on the direct effects of parent wealth on child outcomes.

2 Background

The settlement of Oklahoma has a unique history. Modern day Oklahoma was created from Indian Territory, a place designated for the resettling of American Indian tribes. Beginning with the first land run of 1889, Indian Territory was gradually opened up to outside settlers and a separate Oklahoma Territory was formed in the west. 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. Under the Homestead Act, individuals were allowed to settle plots of 160 acres of public land after filing a claim with the local land office and paying a small fee.[3] If the homesteader stayed on the land and improved it over the next five years, they would then be awarded ownership of the land for a nominal fee and issued a patent (or deed of title) giving them ownership of the land free and clear.[4] Following the land run of 1889, 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, the settlement of Oklahoma occurred quickly within a relatively narrow window of time.

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 passive income from the sale of oil and gas in the form of royalties, typically at the rate of 12.5 percent of production value. Using a back-of-the-envelope calculation based on historical oil prices, the oil discoveries for children in our sample averaged $1.4 million in royalties with a median royalty of $247,524 in 2023 dollars.

[3] Although the original Homestead Act entitled each individual to a 160-acre plot of land, some settlers obtained multiple land patents through cash entries or the Enlarged Homestead Act of 1909. In a small number of cases, a husband and wife received patents for separate plots of land.

[4] Homesteaders were allowed to take ownership sooner than five years by paying a much higher price of $1.25 per acre.

3 Data

We construct a dataset of the children of homesteaders in Oklahoma Territory, combining information on whether oil was discovered on their parent's land and the children's subsequent outcomes in the 1940 U.S. Census. Further details about the data we use are contained in Appendix section **??**.

Our homesteaders are drawn from the Bureau of Land Management (BLM) Federal Land Patents (Bureau of Land Management, U.S. Department of the Interior, 2022), which identify the name, precise location of residence, and patent (deed or title) date for over 170,000 land patents transferred from the federal government to individual homesteaders in Oklahoma Territory.[5] We focus on land patents that were filed between 1889 and 1910 in Oklahoma Territory. After filing for an initial homestead, there were three possible outcomes—prove, abandon, or commute the homestead. To successfully prove a homestead, individuals applied for a patent after at least five years of "settlement and cultivation" (https://www.archives.gov/milestone-documents/homestead-act). Up to 45 percent of initial homestead claims were abandoned and returned to the federal government (Hauck, 2023). Some individuals instead chose to commute their homesteads by purchasing them before the five years were completed. As we explain in Appendix **??**, we focus on proved homesteads for two reasons—to reduce attenuation bias due to out-migration of abandoned homesteads and to eliminate any selection bias from strategic land purchases.

The Public Land Survey System (PLSS) defines a grid of six-mile by six-mile townships, each with 36 one-square mile sections (640 acres). Typically, a homesteader received a quarter-section of land amounting to 160 acres. Thus, a section accommodates four homesteaders and their families, and a township accommodates 136 households.[6] Figure 1 demonstrates that the timing of the first land patents in each section corresponds closely to the timing of land openings (through land runs and land lotteries). The figure also shows that sections were settled relatively quickly and are not clustered on oil fields (see Figure 3).

For each homesteader, we identify whether oil was present on their land by linking the section where they acquired a land patent by 1910 to the locations of post-1910 oil wells. We obtain data on oil wells from the U.S. Geological Survey (USGS), the Oklahoma Geological Survey (OGS), and a private company called WellDatabase. Each dataset provides the locations and drilling dates for oil and gas wells. Figure 4 displays the number of sections where oil was discovered in each year. We count a section as having an oil discovery if at least two of our data sources indicate an oil discovery.[7] If the datasets disagree on the date of discovery, we use the

[5]Most of the land patents in Indian Territory (the eastern portion of present-day Oklahoma) were tribal allotments, which were quite different from the process of homesteading.

[6]Two of the sections were set aside as "School Lands" and, at least initially, could only be leased, not homesteaded.

[7]See the 4 section for more details.

median year. The map in Figure 5 illustrates the variation in the timing and geography of oil discoveries.

The data sources for oil wells do not agree completely on these locations (see Figure 2).[8] Disagreements are most frequent around the edges of apparent oil fields, which may reflect exploratory wells that came up dry and were thus less likely to be recorded in some data sources.[9] Figure 3 labels each section of Oklahoma Territory according to how many data sources indicate the presence of oil in that section. Our labels closely align with the historical map in Figure 2, indicating that we can minimize both false positives and negatives by combining information from OGS, USGS, and WellDatabase.[10] In Section 4 we provide more detail on how we combine data from all three data sources to infer the locations of oil discoveries.

Oil production volumes, available only from the USGS, allow us to impute the size of wealth shocks but do not serve as our primary treatment variable. A major reason for this is that annual production is unavailable prior to 1968. Instead, the USGS data reports cumulative oil production through 1968. Nevertheless, we use production volumes as a measure of treatment intensity.

In order to follow the children of homesteaders over time, we first link homesteaders from the BLM Land Patents to individual census records in the 1900 and 1910 U.S. Census (Ruggles et al. (2024a), Ruggles et al. (2024b), Helgertz et al. (2024)). Census records include personal information that is valuable for record linking, including an individual's first and last name, age, sex, birthplace, residential location, and family members. However, the information in the BLM Land Patents is much more sparse. Federal land patents do not record the age, sex, birthplace, race, occupation, or family members of homesteaders. Rather, they simply record the full name of the homesteader and a detailed description of the land. We develop a strategy to uniquely identify homesteaders in the U.S. Census, despite the sparse demographic information available for linking.

Our sample contains approximately 172,000 land patents, each with a full name and precise geographic location. The goal is to identify a unique census record in 1900 or 1910 corresponding to the homesteader listed on each land patent.[11] A common procedure for linking historical records is to clean first and last names (to account for enumeration and transcription errors),

[8]Figure 2 illustrates differences across data sources compared to a 1931 map of the Oklahoma City oil field. Each dataset traces the contours of the field reasonably well, with minor disagreements around the boundaries of the field.

[9]Historical sources (see Figure 2 and Figure 9 in the appendix) confirm that productive oil fields had a high density of wells, so observing just one or two wells in a section suggests that oil was not discovered. To avoid false positives when assigning the presence of oil, we require five or more wells in a section to designate the section as containing oil in either the OGS or USGS data sources. WellDatabase has approximately 90 percent fewer wells, so we only require the presence of a single well to designate a section as containing oil. WellDatabase is a private company that sells data on production of current wells, so the wells it reports are less likely to be dry, exploratory wells.

[10]See also Figure 9 in the appendix.

[11]This is a many-to-one merging problem because some individuals received multiple land patents.

then to search for unique matches on the basis of names, sex, birthplace, and birth year. In our case, names are the only personally identifiable information in the Federal Land Patents, making it unlikely that we would identify many unique links among the full population (the population of Oklahoma was 400,000 in 1900 and 1.7 million in 1910). Consider the case of Frank Johnson, who acquired a land patent in 1899 located in Oklahoma County (Township 11 North, Range 1 West, Section 9). The 1900 Census contains three adult men named Frank Johnson who resided in Oklahoma County, each of which could plausibly be a match for the homesteader.[12] We need additional information to identify the correct Frank Johnson.

Although the demographic information from the Land Patents is insufficient for linking to the census, there is a key advantage in the Land Patents—information about neighbors. Figure 6 contains a plat map of homesteads in Oklahoma County, with the locations of Frank Johnson and other homesteaders in the same township.[13] Using the names of Frank Johnson's neighbors, we can identify his 1900 Census record due to several of his neighbors being enumerated on the same census sheet. Because census takers traversed neighborhoods on foot, being enumerated on the same or adjacent census sheets is a common measure of geographic proximity.[14] The census link resulting from this procedure is for a Frank Johnson who was born in 1862 in Illinois, and was married to Susie Johnson and living with a 14-year-old son named Frank C. Johnson in 1900.

We develop a simple, three-step algorithm to automate the process of linking tens of thousands of land patents to individual census records, leveraging the critical geographic information encoded in both datasets. In the first step, we limit the set of candidate links for a given homesteader by identifying census records with the same county of residence and a similar name.[15] In the second step, we remove candidate links for individuals who were too young to have qualified for a homestead at the time they received a land patent.[16] This criterion is especially helpful in cases where a parent and child share the same name and live in the same household. After this

[12]The census identifies smaller geographies than counties, but counties are the smallest geographic unit common to both the Federal Land Patents and the census.

[13]This map records the locations of "89ers"—those individuals who claimed a homestead in the 1889 land run and subsequently obtained a land patent. The map can be accessed at `okhistory.org/research/digital/US_Land_Patents_1889.pdf`.

[14]Scholars have used census sheets to measure the effects of social interaction with racial minorities Brown (2021) and to track neighborhoods over time Haws et al. (2025).

[15]Blocking on the county of residence is necessary for our neighborhood-based approach, but it does eliminate candidate links for individuals who left the county between receiving their patent and the census. We require candidate links to have the same first and last name initials across data sources, as well as Jaro-Winkler similarity scores of at least 0.87 for first and last names. Because some nicknames or misspelled names may not match using these criteria, we also include candidate links with similar standardized names. Similar to Price et al. (2021), we tabulate common nicknames using established census links. For instance, "Eliza" has a high Jaro-Winkler score with "Elizabeth," while the nicknames "Bessie" and "Lizzie" do not. The nicknames crosswalk allows us to consider "Bessie" and "Lizzie" as similar to "Elizabeth" as well. Other common nicknames that Jaro-Winkler alone would miss include "Bob" for "Robert" and "Wm" for "William."

[16]Eligible homesteaders included individuals 21 years or older or heads of a family. Other requirements included citizenship or intention to naturalize as well as never having taken arms against the federal government (see Act of May 20, 1862).

second step, many land patents still contain multiple candidate links to census records. In the third step, we use neighbors to choose the census link for each land patent that is most likely to be correct. As illustrated in the case of Frank Johnson, when a land patent could plausibly be linked to multiple census records, we choose the census record with the highest number of neighboring individuals (measured using groups of five adjacent census sheets) with a plausible link to a land patent in the same township. [17]

We apply this three-step process for linking land patents to the 1900 (all patents) and 1910 Census (patents in 1900 or later). When one land patent is linked to a record in both 1900 and 1910, we ensure that the implied census-to-census link is in harmony with the Census Tree (Buckles et al., 2023a). Our linking process results in obtaining a census link for 57.9 percent of land patents in our sample (see Table 1). Although our match rate would likely be higher if we were able to track migrants who moved before the next census, matching these migrant homesteaders is not necessary for our empirical strategy since all such migration decisions were made prior to the discovery of oil.

Once we have linked a homesteader to either the 1900 or 1910 census, we use the Census Tree to find that homesteader's children and follow them forward to 1940. The Census Tree is the largest available set of historical US census links and covers each available decennial census from 1850 to 1940. These links enable researchers to follow individuals and their families across censuses. The links are established by combining machine learning tools with a massive set of census links validated by volunteer genealogists. The Census Tree also includes a much larger set of links for women, which enables us to follow both the sons and daughters of homesteaders. After linking homesteaders to the 1900 or 1910 censuses, we use the Census Tree Project (Buckles et al., 2023b) to identify each homesteader's children. For a given homesteader, we define their children to be anyone who is listed as a child in any census in which the homesteader appears. We then use the Census Tree to follow those children forward to the 1940 Census where we observe their long-run outcomes.

In the end, we compile a dataset of children in 1940 who had a parent who applied for a homestead in Oklahoma. Tables 3 and 4 report summary statistics for our sample of children while Table 2 reports summary statistics for their homesteading parents. Each homesteader in our sample contributes an average of two children to the final sample (see Table 2). There are no significant differences in gender, race, or age across treated and non-treated homesteaders within the same township. Tables 3 and 4 also report raw differences in outcomes for children, 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), human capital (years of education), and marriage market outcomes (ever married, spouse's education). Our empirical strategy involves comparing children

[17] If there is an exact tie, most likely because no neighbors are linked, we favor candidate links with an exact match on the full original name (including middle initials) and discard remaining conflicts.

of homesteaders with oil on their land to children of their neighbors who did not have oil. In our sample, 1,418 sons and 954 daughters had oil while 52,593 sons and 36,347 daughters did not. Restricting to townships with oil, 29,737 children of homesteaders are not treated by oil.[18]

4 Empirical Strategy

We compare the children of neighboring homesteaders with and without oil, assuming that the individuals and the land they reside on are similar. Focusing on townships makes this assumption more believable because geography, climate, and economic characteristics are unlikely to vary greatly at such a local level. Specifically, 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 Tables 1–3). 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 identifying assumption that the presence or absence of oil among early homesteaders was as good as randomly assigned.

4.1 Estimating the effects of oil discovery

To estimate the causal effect of an oil discovery, we compare the children of homesteaders with oil discoveries to their neighbors who never experienced an oil discovery. However, there are good reasons to worry that our OLS estimates are attenuated due to measurement error in whether or not oil is present. 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 homesteaders 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 with the wrong date or by 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 themselves. Ferrara et al. (2024) observe that this is a common problem when using historical data and provide guidance on dealing with the attenuation bias this measurement error creates. They suggest three solutions which all depend on having multiple noisy measures of the right hand side variable.

Ferrara et al. (2024)'s preferred solution is to generate an "agreement sample." The agreement sample includes only observations where the noisy measures agree on whether the observation was treated or not. They find that limiting the sample to the agreement sample can eliminate much of the attenuation bias. This method also allows us to estimate nonlinear and interactive effects, which are important in our application. We define the agreement sample to

[18]This statistic is not part of Tables 3 and 4.

include children for whom (i) all three datasets agree that oil was not discovered before 1940, or (ii) at least two datasets agree that oil was discovered before 1940. By restricting our analysis to the agreement sample, we eliminate much of the measurement error in treatment status.

Ferrara et al. (2024) also propose two additional solutions to measurement error. First, they propose bounding the treatment effect with OLS and IV. The OLS estimate is attenuated and therefore biased toward zero while the IV estimate (using one noisy measure as an instrument for the other) is biased away from zero. Thus, OLS and IV estimates can provide bounds on the true treatment effect. Second, they propose constructing a nonlinear function of the OLS and IV estimates that is a consistent estimator of the treatment effect. They suggest the bias-corrected estimator $\delta^{BC} \equiv \sqrt{\delta^{OLS} \times \delta^{IV}}$. One drawback of this approach is that the estimator is only defined if the OLS and IV estimates have the same sign.

Our primary estimates come from Ferrara et al. (2024)'s agreement sample approach. Define W_j^{OGS}, W_j^{USGS}, and W_j^{WD} to be the number of wells observed on homesteader j's land in the OGS, USGS, and WellDatabase datasets. Then define $Z_j^{OGS} \equiv \mathbb{1}\{W_j^{OGS} \geq 5\}$ and $Z_j^{USGS} \equiv \mathbb{1}\{W_j^{USGS} \geq 5\}$ to be dummies equal to one if the OGS and USGS well counts are greater than or equal to five. We define $Z_j^{WD} \equiv \mathbb{1}\{W_j^{WD} \geq 1\}$ for the WellDatabase dataset because, as discussed above, it contains far fewer wells, probably because it omits many dry wells. We limit the sample to the agreement sample ($Z^{OGS} + Z^{USGS} + Z^{WD} \geq 2$ or $Z^{OGS} + Z^{USGS} + Z^{WD} = 0$) and estimate the following regression

$$y_{ijt} = \delta_0 + \delta^{AS} \mathbb{1}\{Z_j^{OGS} + Z_j^{USGS} + Z_j^{WD} \geq 2\} + X_i \cdot \beta + a_t + \epsilon_{ijt} \tag{1}$$

where y_{ij} is an outcome in 1940 for child i of homesteader j living in township t, the X_i are child covariates,[19] and a_t is a set of township fixed effects.[20] The parameter δ^{AS} represents the effect of an oil discovery on y for the children of homesteaders in 1940 estimated from the agreement sample. By including township fixed effects, our estimator compares the children of homesteaders with oil to the children of their neighbors who did not have oil. When the outcome is binary, we fit a logit model rather than OLS. In Appendix B we also report estimates using Ferrara et al. (2024)'s bounding and bias-correction approaches. The findings are qualitatively similar to those reported in Section 5 below, albeit without the ability to interact treatment with child characteristics.

4.2 Interactions with age

When estimating the effects of parental wealth on children, we interact our treatment dummy with either the parent's age in 1940, the child's age in 1940, or the child's age at discovery. When

[19] We include gender, race, and age dummies as controls.

[20] 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." A typical section contains four homesteaders with 160 acres of land each, so each township contains roughly 144 homesteaders.

estimating effects on proxies for child wealth, we interact treatment with a dummy for whether the parent would have been over 70 years old in 1940. Our rationale is that these parents are more likely to have passed away and (potentially) bequeathed wealth to their children. For other outcomes, the age of the child is key. For instance, we could only hope to detect an effect of oil discovery on school enrollment for children who were still school-age in 1940. By the same token, we would only expect to find an effect on educational attainment for children were young when the discovery occurred *and* were old enough in 1940 to have completed their education. And we would expect to see an effect on labor market and marriage market outcomes only for children who had reached adulthood by 1940. Thus, when an outcome depends on the age of the child, we add to equation (1) an interaction between treatment and either age at treatment (in the case of educational attainment, labor supply, earnings, and migration) or age in 1940 (in the case of school enrollment and marriage).[21]

4.3 Using production volumes to estimate treatment intensity

We assign treatment status based on the number of wells present in each of the three datasets, but the magnitude of these oil wealth shocks depends on the actual amount of oil being produced, which could vary considerably. Unfortunately, among our three datasets only the USGS dataset contains information on oil production, and even that is cumulative production volumes in 1968. Although not ideal, we still use the USGS production data to identify children whose parents experienced especially large wealth shocks. According to the USGS, a small number of our treated children live in sections where zero oil production was recorded. Among those with positive reported production, the median is 76,335 barrels which corresponds to $247,524 of oil wealth in 2023 dollars (see Table **??**). To focus on children with "large" oil discoveries, we drop treated children with zero recorded production and those with production below the median of 76,335 barrels. Among these high-intensity treated children, the median wealth shock was $1.1 million while the average wealth shock exceeded $2.7 million.

5 Results

5.1 Children's wealth

Although the census does not directly measure wealth, it does measure several proxies for wealth: owning a home, whether the home value exceeds $5,000, and whether the child had more than $50 in nonwage income (a proxy for receiving oil royalties).[22] In Tables 6 and 7 we regress these measures of wealth on a dummy for treatment and an interaction with a dummy for whether the child's parent (the parent who discovered oil), would be under 70 years old in 1940. We

[21] We also include interactions with a dummy variable for the child being under 18 years of age.

[22] Although 55 percent of the sons in our sample own a home, only 6 percent have a home in excess of $5,000. 49 percent have nonwage income in excess of $50.

find that oil discoveries raise the wealth of children with older parents, but not those with younger parents. Moreover, the effects are largest for the high-intensity children—those with oil discoveries above the median. We do not find a significant difference in the effects for sons vs daughters. These results suggest that intergenerational wealth is transferred primarily through bequests from parents to children, rather than through *inter vivos* transfers and that these bequests accrued to both sons and daughters.

5.2 School enrollment and educational attainment

Many of the children in our sample were already adults by the time their parents discovered oil. For these children, we would not expect to see an effect on school enrollment or educational attainment. Among those who were young at the time of discovery, their ages in 1940 vary considerably. For school enrollment, we focus on children who are still young enough to be enrolled in school, so we limit the sample to those between ages 14–24 in 1940.[23] In contrast, for educational attainment we focus on children who are old enough to have completed their education by limiting the sample to those at least 20 years old in 1940.[24] These age restrictions also mean that we are using different birth cohorts in each estimate. For school enrollment, we focus on cohorts born between 1916–1936 (14–24 years old in 1940) who had oil discovered in their townships anytime before 1940 while for educational attainment we focus on cohorts born before 1920 (20+ years old in 1940) who had oil discovered in their townships between their birth year and 1940. Thus, the school enrollment sample tends to include younger cohorts from later discoveries while the educational attainment sample tends to include older cohorts from earlier discoveries.

For school enrollment, we fit a logit model and interact our treatment dummy with the child's age in 1940. Figure 7 plots the predicted probability of enrollment by age in 1940. Between ages 14–24, we find that children with oil are more likely to be enrolled in school than their neighbors without oil ($p < 0.01$). For instance, 94 percent of 17-year-olds with oil are enrolled in school compared with only 71 percent of their neighbors. These higher enrollment rates imply 5.8 cumulative years of schooling between 14–24 for children with oil compared with only 5.0 years for their neighbors, nearly a full year of additional schooling.

For educational attainment, we interact our treatment dummy with the child's age when oil was discovered in his township and a dummy for whether the child was under 18 when discovery occurred. We also estimate these regressions separately for sons and daughters. Tables 8 and 9 report regression coefficients while Tables 10 and 11 report predicted marginal effects of treatment at various ages. As expected, we find small and insignificant effects for children who were already adults at the time of discovery. But for those who were age 5 when discovery

[23] Very few of the children in our sample are less than 14 years old. Among those that are, nearly all of them are enrolled in school.

[24] College attendance is relatively rare in our sample, so most children have completed their education by age 20.

occurred, their educational attainment rises by 0.7 years for sons and by 0.9 years for daughters. Unfortunately, these estimates are noisy and are only significant for sons at the 10 percent level, largely because only a small number of children in our sample were both very young at the time of discovery *and* over 20 years old in 1940. Still the effects we find for young children are consistent with our estimates of the effect on school enrollment.

5.3 Labor supply, earnings, and migration

If oil discoveries increase schooling, it seems reasonable to suspect they may also affect labor market outcomes. However, we find no difference between children with oil and their neighbors in the probability of being in the labor force, weekly hours worked, or wage income. However, we do find suggestive evidence that sons and daughters who were young at the time of discovery are more likely to work in white collar occupations and that sons are more likely to live in a city.

5.4 Marriage

As with education, we would expect the effects of oil discoveries on marriage to be concentrated around certain ages. By age 40, most of the children in our sample have been married, so we focus on children who were between 18–40 years old in 1940 and who had a discovery in their township by the time they turned 18. We estimate a logit model for the probability of ever being married separately for sons and daughters and interact age in 1940 with a dummy for treatment. Figure 8 plots predicted probabilities of ever being married by age. We find a small negative, but statistically insignificant effect, on the age of marriage for sons. For daughters though, oil discoveries appear to delay marriage considerably, by 1.4 years at the median and by 3.2 years at the 75th percentile. Nevertheless, by age 40 nearly all daughters have married, regardless of parental wealth. What's more, it appears that daughters with oil who were young at the time of discovery also tend to marry more educated spouses—1.7 more years of schooling for daughters that were zero years old at discovery (see Table 11). These results are consistent with a story where parental wealth gives daughters greater freedom to delay marriage in search of a better match.

6 Conclusion

We leverage the quasi-random discovery of oil 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 oil discoveries increase several proxies for children's wealth, but only among children whose parents are most likely to

be deceased by 1940, which suggests that oil wealth is transferred primarily through bequests from parents to children, rather than through *inter vivos* transfers. We also find evidence that wealth shocks affect outcomes beyond children's monetary wealth. Oil discoveries caused children between ages 14–24 to enroll in school at higher rates in 1940. Although we find no effect on labor market participation, wages, or hours worked, we do find suggestive evidence that children who were young at the time of discovery are more likely to work in white collar occupations and that sons are more likely to live in a city. We find that oil discoveries cause daughters to delay marriage and to ultimately marry a more educated spouse, suggesting that parental wealth gives daughters greater freedom to delay marriage in search of a better match.

Our results inform a large literature that has focused on estimating the correlation between parent income or wealth and child outcomes. We are one of only a few papers that has used natural experiments to estimate the *causal* effects of parent wealth on children's outcomes (Ager et al., 2021; 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, marriage, and measures of wealth.

In addition, our paper builds on the Census Tree project (Buckles et al., 2023b) 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

Act of May 20, 1862. *(Homestead Act), Public Law 37-64 (12 STAT 392); 5/20/1862; Enrolled Acts and Resolutions of Congress, 1789 - 2011; General Records of the United States Government, Record Group 11; National Archives Building*. Washington, DC.

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.

Aymeric Bellon, J Anthony Cookson, Erik P Gilje, and Rawley Z Heimer. Personal wealth, self-employment, and business ownership. *The Review of Financial Studies*, 34(8):3935–3975, 04 2021. ISSN 0893-9454. doi: 10.1093/rfs/hhab044. URL `https://doi.org/10.1093/rfs/hhab044`.

Dan Black, Terra McKinnish, and Seth Sanders. The economic impact of the coal boom and bust. *The Economic Journal*, 115:449–476, 4 2005a. URL `https://academic.oup.com/ej/article/115/503/449/5086021`.

Dan A Black, Terra G McKinnish, and Seth G Sanders. Tight labor markets and the demand for education: Evidence from the coal boom and bust. *Industrial and Labor Relations Review*, 59, 10 2005b. URL `www.jstor.org`.

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.

Jason P. Brown. Response of consumer debt to income shocks: The case of energy booms and busts. *Journal of Money, Credit and Banking*, 53:1629–1675, 10 2021. ISSN 15384616. doi: 10.1111/jmcb.12842.

Kasey Buckles, Adrian Haws, Joseph Price, and Haley E.B. Wilbert. Breakthroughs in historical record linking using genealogy data: The census tree project. Working Paper 31671, National Bureau of Economic Research, September 2023a.

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*, 2023b.

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

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.

Bureau of Land Management, U.S. Department of the Interior. General Land Office (GLO) Records. `https://glorecords.blm.gov`, 2022. Accessed: 2022-12-02.

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.

J. Anthony Cookson, Erik P. Gilje, and Rawley Z. Heimer. Shale shocked: Cash windfalls and household debt repayment. *Journal of Financial Economics*, 146(3):905–931, 2022. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2022.09.008. URL `https://www.sciencedirect.com/science/article/pii/S0304405X22002069`.

James Cust and Steven Poelhekke. The local economic impacts of natural resource extraction, 10 2015. ISSN 19411359.

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

Andreas Ferrara, Joung Yeob Ha, and Randall Walsh. Using digitized newspapers to address measurement error in historical data. *The Journal of Economic History*, 84(1):271–306, 2024. doi: 10.1017/S0022050723000530.

James Feyrer, Erin T. Mansur, and Bruce Sacerdote. Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review*, 107:1313–1334, 4 2017. ISSN 00028282. doi: 10.1257/aer.20151326.

Kenny A Franks. *The Oklahoma Petroleum Industry*. University of Oklahoma Press, Norman, 1980.

Katherine Hauck. Individual level and farm level homesteading characteristics: A descriptive overview. Working Paper 4566300, SSRN, 2023.

Adrian Haws, David R. Just, and Joseph Price. Who (actually) gets the farm? Intergenerational farm succession in the United States. *American Journal of Agricultural Economics*, 107(1): 3–26, 2025. doi: https://doi.org/10.1111/ajae.12493.

Jonas Helgertz, Nesile Ozder, Steven Ruggles, John Robert Warren, Catherine A. Fitch, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. IPUMS Multigenerational Longitudinal Panel: Version 1.2 [dataset]. Minneapolis, MN: IPUMS, 2024. URL `https://doi.org/10.18128/D016.V1.2`.

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

Amir Mousavi and Jeremy Edmund Clark. The effects of natural resources on human capital accumulation: A literature survey. *Journal of Economic Surveys*, 35:1073–1117, 9 2021. ISSN 14676419. doi: 10.1111/joes.12441.

Lucija Muehlenbachs, Elisheba Spiller, and Christopher Timmins. The housing market impacts of shale gas development. *American Economic Review*, 105:3633–3659, 12 2015. ISSN 00028282. doi: 10.1257/aer.20140079.

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.

Joseph Price, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. Combining family history and machine learning to link historical records: The census tree data set. *Explorations in Economic History*, 80, 4 2021.

Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024a. URL `https://doi.org/10.18128/D010.V15.0`.

Steven Ruggles, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. IPUMS ancestry full count data: Version 4.0 [dataset]. Minneapolis, MN: IPUMS, 2024b. URL `https://doi.org/10.18128/D014.V4.0`.

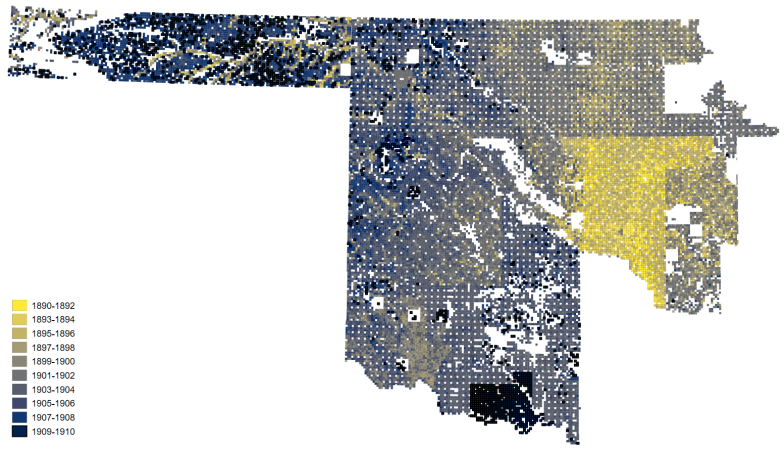
Xi Song, Catherine G Massey, Karen A Rolf, Joseph P Ferrie, Jonathan L Rothbaum, and Yu Xie. Long-term Decline in Intergenerational Mobility in the United States Since the 1850s. *Proceedings of the National Academy of Sciences*, 111:251–258, 10 2019.

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

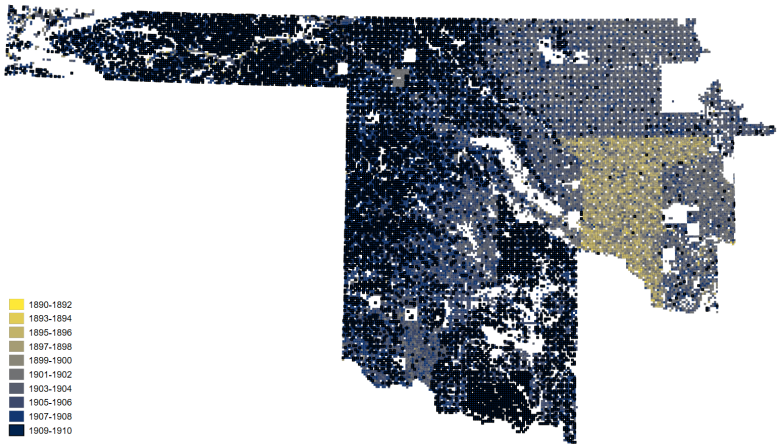
7 Tables and Figures

Figure 1: Oklahoma Territory - Land Patent Years

(a) First year of land patents within a section



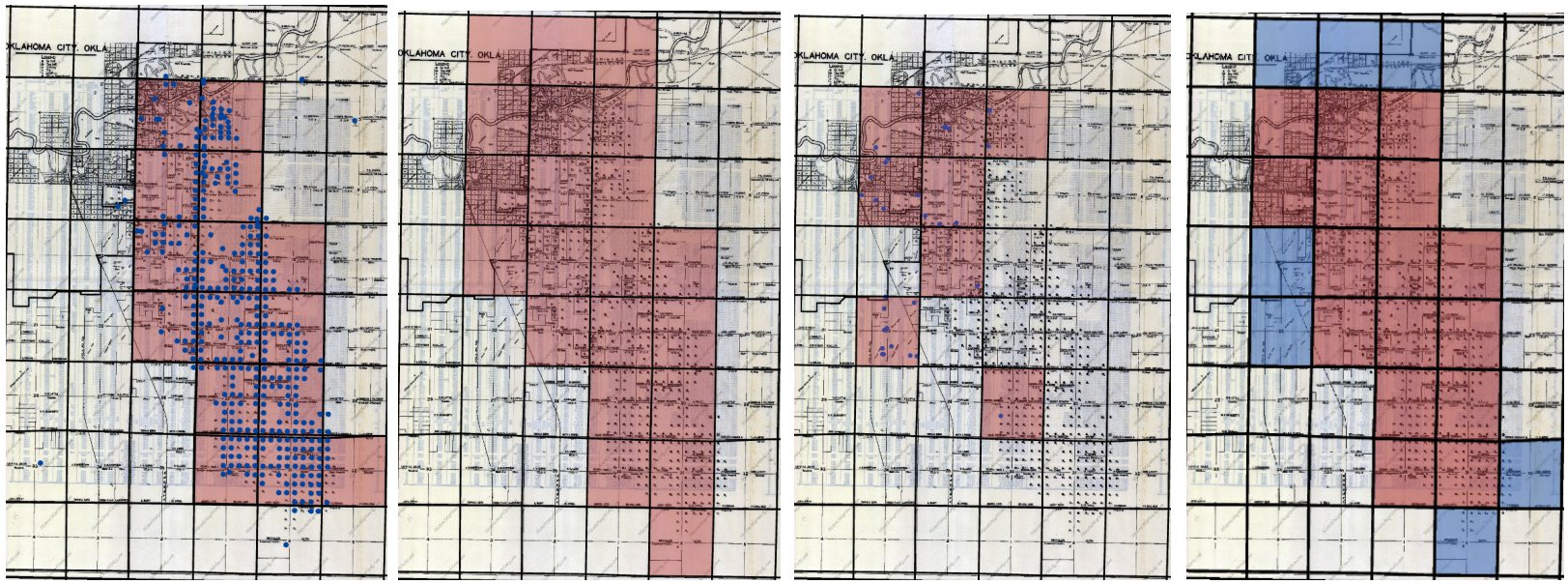
(b) Last year of land patents within a section



Notes: Each section in Oklahoma Territory shows the first year (panel a) and last year (panel b) a homesteading land patent was issued between 1890 and 1910 from the Bureau of Land Management General Land Office Records.

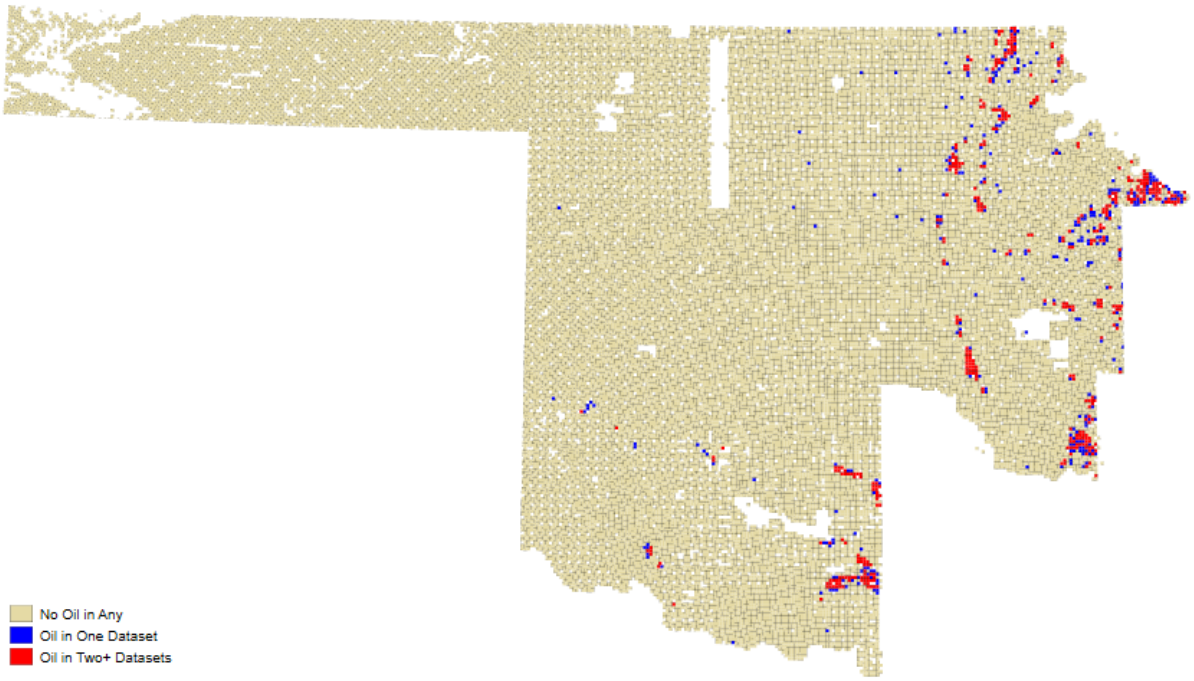
Figure 2: Oklahoma City Oil Field Map by 1930

(a) OK Geol. Survey (b) U.S. Geol. Survey (c) WellDatabase (d) All Oil



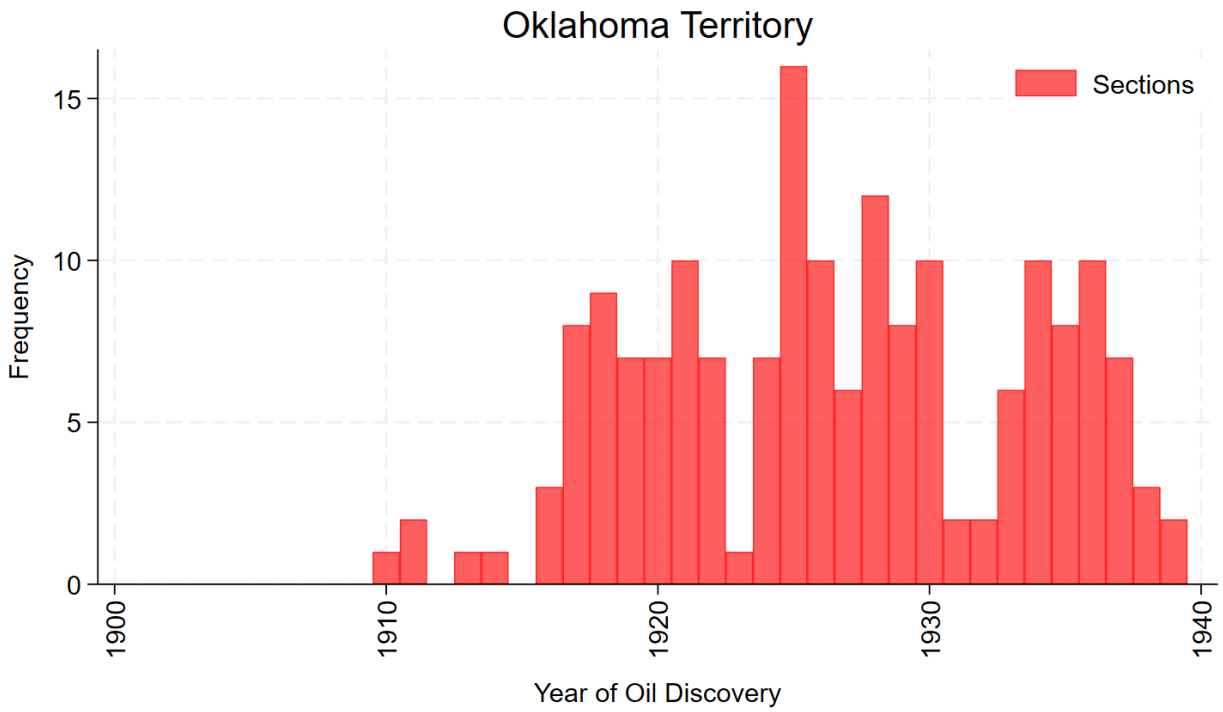
Notes: This historical map of the Oklahoma City Field from *The Oil Weekly* (August 1, 1930) plots PLSS sections and the locations of oil wells in black. Red squares in (a) represent five or more wells drilled within the section before 1931 according to OGS. Red squares in (b) represent five or more wells drilled before 1931 according to USGS. Red squares in (c) represent one or more wells drilled before 1931 according to WellDatabase. Blue dots indicate oil wells in (a) and (c). Blue squares in (d) represent sections where only one dataset shows oil is present. Red squares represent sections where two or more datasets agree oil is present.

Figure 3: Sections in Oklahoma Territory with Oil



Notes: Tan areas represent sections where homesteading land patents are recorded between 1890-1910 and no oil is present before 1940 in all three oil datasets. Blue areas represent sections where homesteading land patents are recorded between 1890-1910 and oil is present before 1940 in only one of the oil datasets. Red areas represent sections where homesteading land patents are recorded between 1890-1910 and oil is present before 1940 in two or three datasets. White areas represent sections where sale or homestead land patents are not recorded between 1890-1910 due to the land being reserved for school funds, Indian allotments, or government land.

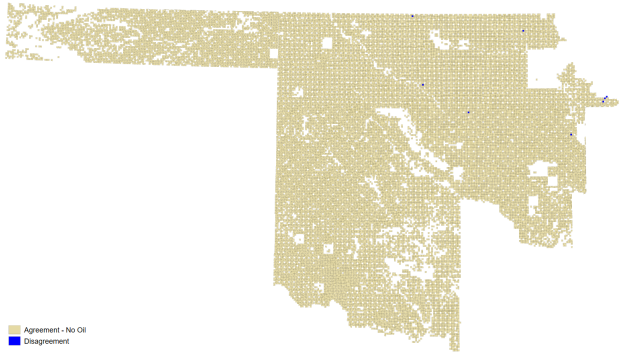
Figure 4: Year of Oil Discovery in Homesteaded Oklahoma Territory



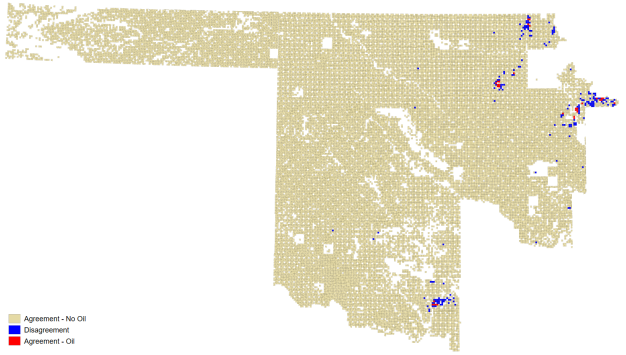
Notes: This figure plots the number of sections where oil was discovered in each year within homesteaded Oklahoma Territory. We define an oil discovery to occur if two or more datasets indicate the presence of oil in the section. If the oil discovery year differs by dataset, we use the median discovery year.

Figure 5: Oil Discoveries In Oklahoma Territory by Decade

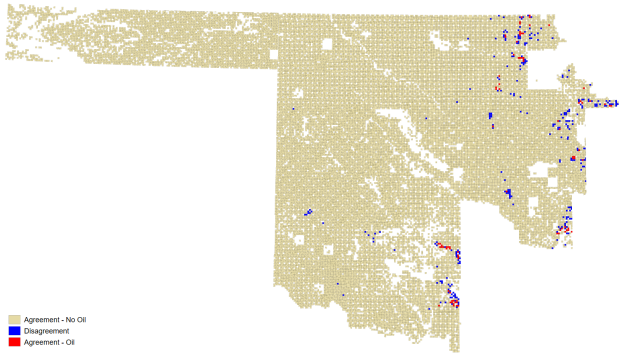
(a) Oil Before 1907



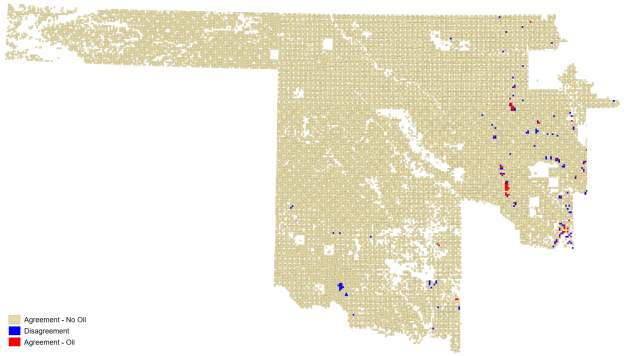
(b) Oil Between 1907-1919



(c) Oil Between 1920-1929



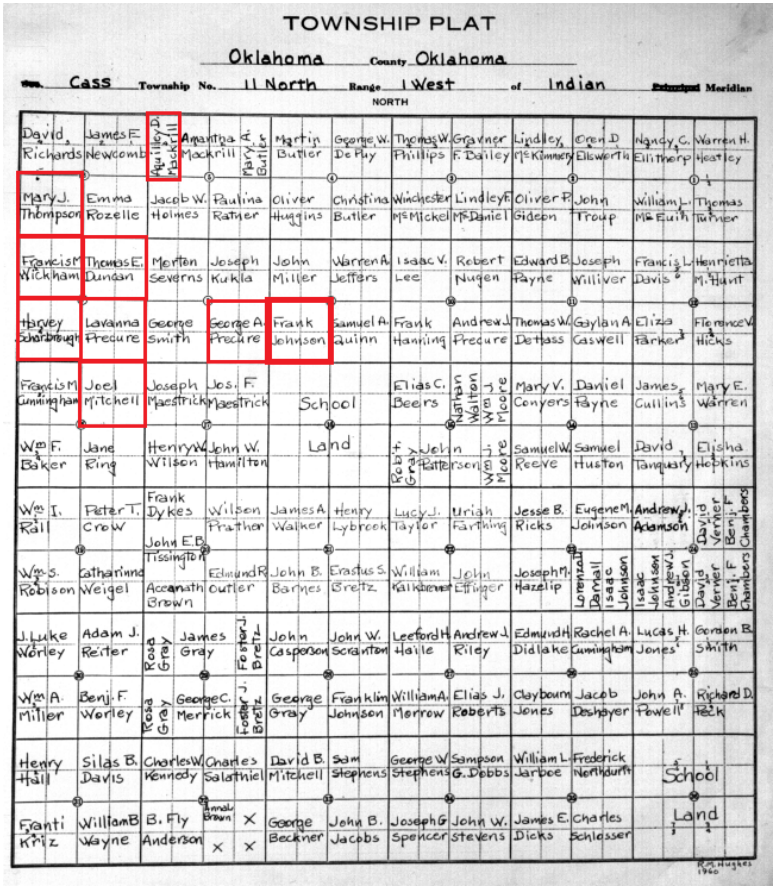
(d) Oil Between 1930-1939



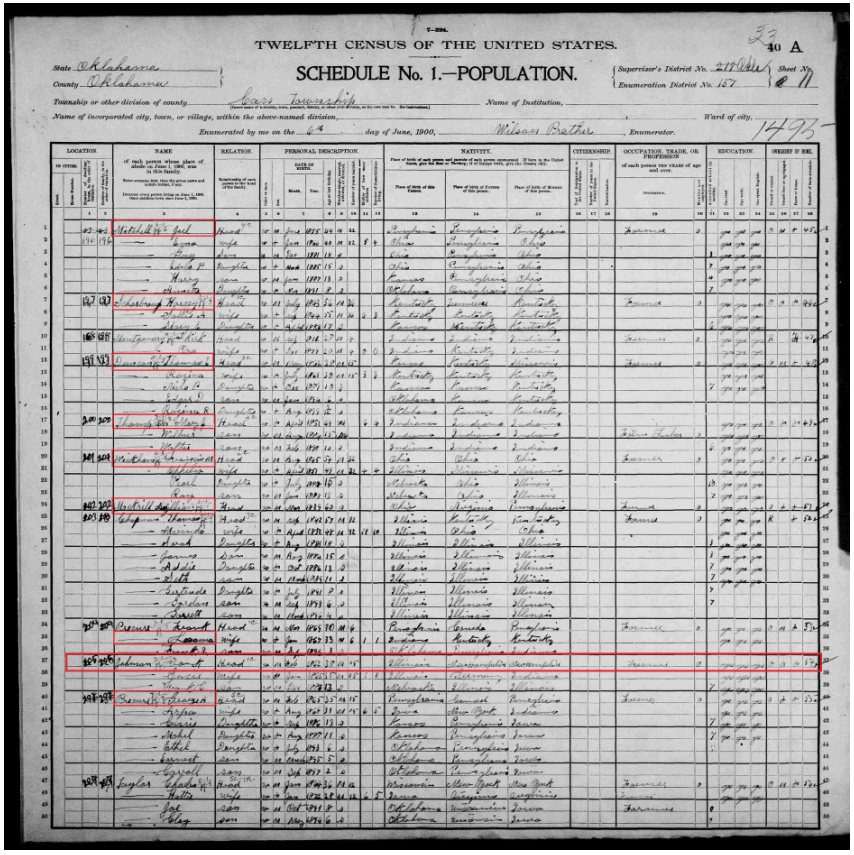
Notes: Tan areas represent sections where a homesteading land patent is recorded between 1890-1910 and no oil is present in the years specified for any dataset. Blue areas represent sections where a homesteading land patent is recorded between 1890-1910 and oil is present in only one dataset. Red areas represent sections where a homesteading land patent is recorded between 1890-1910 and oil is present in two or more datasets. If the oil discovery year differs by dataset, we use the median discovery year. White areas represent sections where no homesteading land patents are recorded between 1890-1910 due to the land being reserved for school funds, Indian allotments, or government land.

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

(a) Plat map

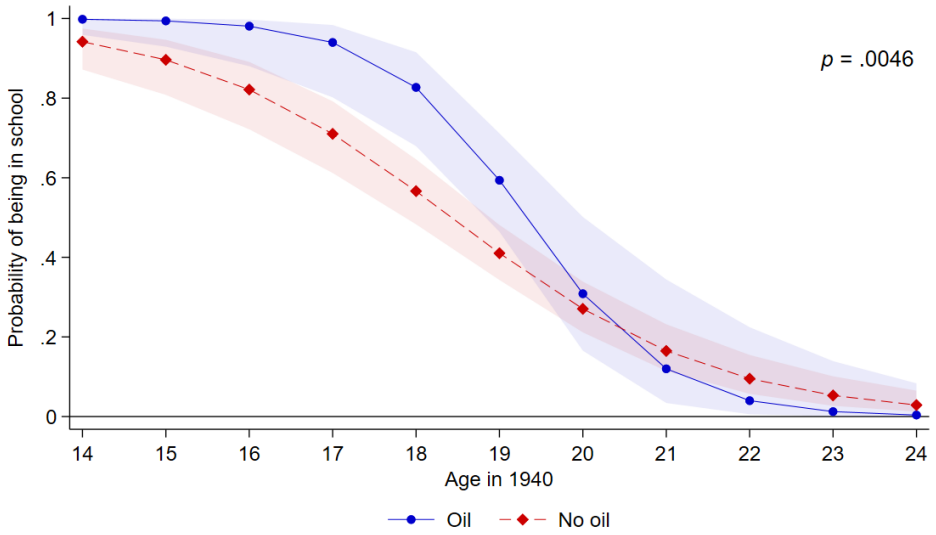


(b) Census page



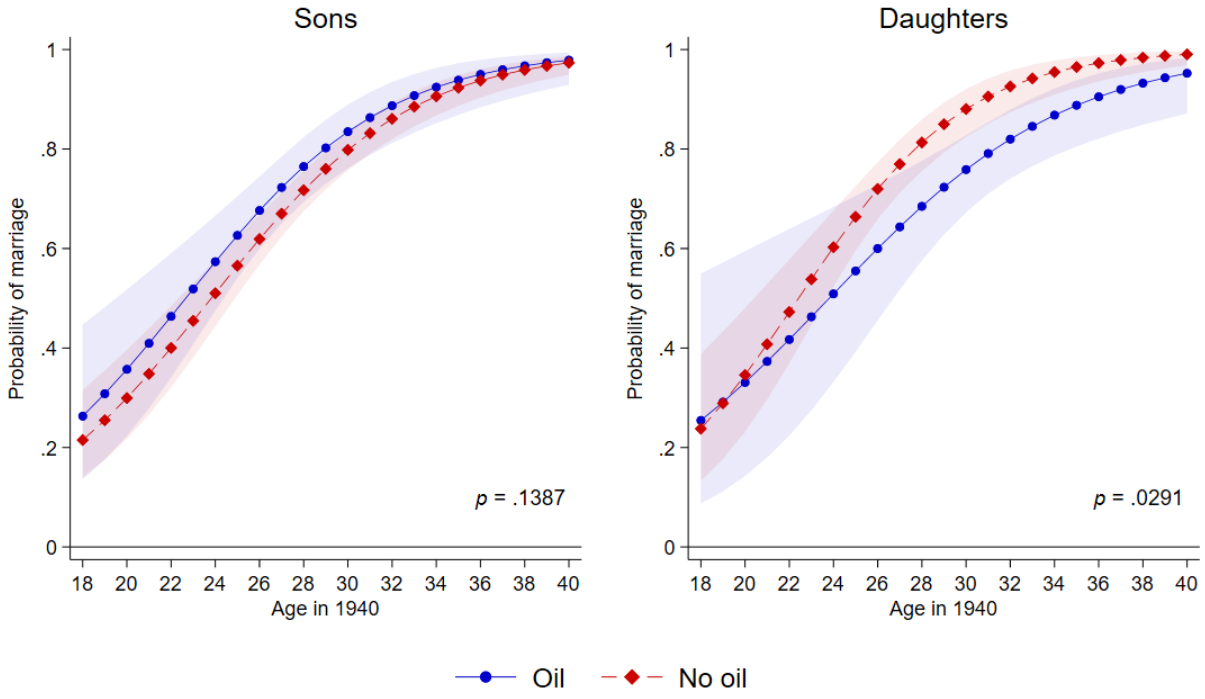
Notes: Panel (a) shows the location of Frank Johnson and several neighbors in a township plat map in Oklahoma County, available from okhistory.org. Panel (b) shows where these neighbors can be co-located on a page of the 1900 U.S. Census in Oklahoma County, downloaded from FamilySearch.org.

Figure 7: Predicted probability of being in school in 1940, by parental oil exposure



Notes: This figure plots predicted probabilities and 95% confidence intervals from a logit regression estimated on the agreement sample. The outcome is a binary indicator for being enrolled in school in 1940. We interact an indicator for oil discovery with a linear term for the child's age in 1940. We also include binary control variables for sex and race. We include township fixed effects and restrict the sample to children from townships with oil discoveries who were between ages 14–24 in 1940. Standard errors are clustered at the section level, and we report the p-value from a joint hypothesis test that the treatment indicator and its interactions are all null.

Figure 8: Predicted probability of ever marrying by 1940, by parental oil exposure



Notes: This figure plots predicted probabilities and 95% confidence intervals from a logit regression estimated on the agreement sample. The outcome is a binary indicator for having married by 1940. We interact an indicator for oil discovery with a linear term for the child's age in 1940. We also include binary control variables for race and estimate the model separately for sons and daughters. We include township fixed effects and restrict the sample to children from townships with oil discoveries by age 18 who were between ages 18–40 in 1940. Standard errors are clustered at the section level, and we report the p-value from a joint hypothesis test that the treatment indicator and its interactions are all null.

Table 1

Summary statistics by linkage to 1900 or 1910 census, BLM land patents

	Linked	Unlinked	Within township difference	Unconditional difference
Oil in section	0.02	0.01	0.00	0.00***
	[0.13]	[0.12]	(0.00)	(0.00)
Year of entry	1926.07	1926.96	-0.39***	-0.89***
	[7.88]	[7.91]	(0.11)	(0.15)
Cash entry	0.20	0.41	-0.16***	-0.21***
	[0.40]	[0.49]	(0.00)	(0.00)
Has middle initial	0.75	0.71	0.04***	0.03***
	[0.43]	[0.45]	(0.00)	(0.00)
Observations	99,780	72,679		

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

Notes: Mean characteristics of BLM land patents, with standard deviations in brackets. Differences are computed as coefficients from a series of regressions with each characteristic as a dependent variable and an indicator for being linked to a census as the key independent variable. Heteroskedasticity-robust standard errors are in parentheses. Within-township differences are from regressions with township fixed effects.

Table 2

Summary statistics, linked homesteaders

	Oil in section	No oil	Within township difference	Unconditional difference
BLM land patents				
Median year of entry	1902.18	1903.61	-0.02	-1.44***
	[3.30]	[3.89]	(0.08)	(0.14)
Any cash entry	0.17	0.20	-0.02*	-0.03**
	[0.38]	[0.40]	(0.01)	(0.01)
Total land patents	1.57	1.71	0.04	-0.14***
	[1.07]	[2.97]	(0.04)	(0.04)
Earliest linked census (1900 or 1910)				
Male	0.92	0.90	0.00	0.02*
	[0.27]	[0.30]	(0.01)	(0.01)
Age in 1900	41.30	38.81	-0.48	2.48***
	[12.73]	[13.90]	(0.42)	(0.37)
Black	0.03	0.03	0.00	0.00
	[0.16]	[0.17]	(0.01)	(0.01)
Indigenous	0.00	0.00	0.00	0.00
	[0.05]	[0.02]	(0.00)	(0.00)
On farm	0.74	0.80	-0.01	-0.06***
	[0.44]	[0.40]	(0.01)	(0.01)
Farmer	0.74	0.67	0.01	0.06***
	[0.44]	[0.47]	(0.02)	(0.01)
Home	0.84	0.80	0.01	0.04***
	[0.37]	[0.40]	(0.01)	(0.01)
Married	0.77	0.72	0.00	0.06***
	[0.42]	[0.45]	(0.02)	(0.01)
Children in any census				
Any children	0.69	0.59	0.00	0.09***
	[0.46]	[0.49]	(0.01)	(0.01)
Number of children linked to 1940	2.01	1.74	0.12*	0.27***
	[2.30]	[2.27]	(0.08)	(0.07)
Observations	1,298	56,367		

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

Notes: Characteristics of individuals in the census who are linked to land patents. Mean values are reported, with standard deviations in brackets. Differences are computed as coefficients from a series of regressions with each characteristic as a dependent variable and an indicator for oil being present in the section as the key independent variable. Standard errors are clustered at the section level and shown in parentheses. Within-township differences are from regressions with township fixed effects.

Table 3

Summary statistics in 1940 census, sons

	Oil in section	Oil in section -- high production	No oil	Difference (I - III)		Difference (II - III)	
	(I)	(II)	(III)	Within township	Unconditional	Within township	Unconditional
Characteristics							
Age	45.29	45.25	44.00	-1.26**	1.29**	-1.31*	1.25*
	[13.35]	[12.87]	[14.02]	(0.64)	(0.57)	(0.76)	(0.74)
Age at township discovery	30.98	32.93	35.81	-1.58**	-4.83***	-1.77**	-2.88***
	[14.88]	[14.31]	[15.64]	(0.65)	(0.80)	(0.76)	(0.98)
Black	0.02	0.02	0.02	0.00	0.00	-0.01	0.00
	[0.14]	[0.12]	[0.14]	(0.01)	(0.01)	(0.01)	(0.01)
Indigenous	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	[0.06]	[0.08]	[0.05]	(0.00)	(0.00)	(0.00)	(0.00)
Age of youngest parent	81.35	81.10	80.03	-1.63**	1.32**	-2.02***	1.07
	[11.72]	[11.11]	[12.70]	(0.63)	(0.57)	(0.74)	(0.72)
Outcomes							
In school	0.03	0.02	0.03		-0.01		-0.01***
	[0.16]	[0.13]	[0.18]		(0.01)		(0.01)
Ever married	0.89	0.89	0.85		0.03***		0.04***
	[0.32]	[0.31]	[0.36]		(0.01)		(0.01)
Years of education	8.82	8.83	8.82		0.00		0.01
	[3.42]	[3.43]	[3.26]		(0.12)		(0.17)
Hours worked	32.59	33.51	34.30		-1.71**		-0.79
	[25.80]	[25.38]	[25.09]		(0.78)		(1.08)
Wage income	698.94	674.38	593.70		105.25***		80.69*
	[942.78]	[946.13]	[857.75]		(30.22)		(42.90)
Spouse's years of education	9.41	9.57	9.47		-0.06		0.10
	[2.89]	[2.89]	[2.91]		(0.11)		(0.14)
In labor force	0.86	0.87	0.88		-0.02**		-0.01
	[0.35]	[0.34]	[0.32]		(0.01)		(0.01)
White collar occupation	0.22	0.25	0.21		0.02		0.04**
	[0.42]	[0.43]	[0.40]		(0.01)		(0.02)
Urban	0.39	0.41	0.37		0.03		0.04**
	[0.49]	[0.49]	[0.48]		(0.02)		(0.02)
Home	0.54	0.56	0.51		0.03**		0.04**
	[0.50]	[0.50]	[0.50]		(0.01)		(0.02)
Home value >$5000	0.06	0.08	0.04		0.02**		0.04***
	[0.24]	[0.27]	[0.20]		(0.01)		(0.01)
Nonwage income >$50	0.46	0.50	0.47		-0.01		0.02
	[0.50]	[0.50]	[0.50]		(0.02)		(0.02)
Observations	1,418	704	52,593				

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

Notes: Characteristics of sons of patent holders in the 1940 census. Mean values are reported, with standard deviations in brackets. Differences are computed as coefficients from a series of regressions with each characteristic as a dependent variable and an indicator for oil being present in the section (or for high production oil) as the key independent variable. Standard errors are clustered at the section level and shown in parentheses. Within-township differences are from regressions with township fixed effects.

Table 4

Summary statistics in 1940 census, daughters

	Oil in section	Oil in section -- high production	No oil	Difference (I - III)		Difference (II - III)	
	(I)	(II)	(III)	Within township	Unconditional	Within township	Unconditional
Characteristics							
Age	43.59	43.16	43.31	-1.99***	0.28	-2.21**	-0.15
	[12.70]	[11.87]	[13.93]	(0.73)	(0.63)	(0.88)	(0.74)
Age at township discovery	28.79	30.01	34.36	-2.61***	-5.57***	-3.04***	-4.36***
	[14.33]	[13.57]	[15.21]	(0.73)	(0.90)	(0.85)	(1.04)
Black	0.01	0.01	0.01	0.00	0.00	0.01	0.00
	[0.10]	[0.10]	[0.08]	(0.01)	(0.00)	(0.01)	(0.01)
Indigenous	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[0.03]	[0.05]	[0.04]	(0.00)	(0.00)	(0.00)	(0.00)
Age of youngest parent	79.82	79.33	79.63	-2.31***	0.19	-2.73***	-0.30
	[11.24]	[10.69]	[12.44]	(0.71)	(0.60)	(0.88)	(0.81)
Outcomes							
In School	0.04	0.03	0.04		0.00		-0.01
	[0.19]	[0.17]	[0.19]		(0.01)		(0.01)
Ever married	0.88	0.87	0.87		0.01		-0.01
	[0.33]	[0.34]	[0.33]		(0.01)		(0.02)
Years of education	9.52	9.52	9.47		0.05		0.05
	[3.18]	[3.17]	[3.09]		(0.14)		(0.20)
Hours worked	5.93	6.78	7.43		-1.51***		-0.65
	[15.83]	[16.70]	[17.62]		(0.54)		(0.80)
Wage income	97.50	108.90	106.57		-9.07		2.33
	[331.32]	[319.50]	[332.70]		(12.21)		(16.30)
Spouse's years of education	8.92	8.89	8.77		0.16		0.12
	[3.33]	[3.28]	[3.19]		(0.15)		(0.21)
In labor force	0.18	0.19	0.19		-0.02		0.00
	[0.38]	[0.40]	[0.40]		(0.01)		(0.02)
White collar occupation	0.10	0.13	0.10		0.00		0.03*
	[0.30]	[0.33]	[0.30]		(0.01)		(0.02)
Urban	0.43	0.44	0.39		0.04**		0.05*
	[0.50]	[0.50]	[0.49]		(0.02)		(0.03)
Home	0.57	0.59	0.56		0.01		0.03
	[0.50]	[0.49]	[0.50]		(0.02)		(0.02)
Home value >$5000	0.08	0.11	0.05		0.03***		0.05***
	[0.28]	[0.31]	[0.23]		(0.01)		(0.02)
Nonwage income >$50	0.18	0.19	0.17		0.01		0.02
	[0.38]	[0.39]	[0.37]		(0.01)		(0.02)
Observations	954	480	36,347				

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

Notes: Characteristics of daughters of patent holders in the 1940 census. Mean values are reported, with standard deviations in brackets. Differences are computed as coefficients from a series of regressions with each characteristic as a dependent variable and an indicator for oil being present in the section (or for high production oil) as the key independent variable. Standard errors are clustered at the section level and shown in parentheses. Within-township differences are from regressions with township fixed effects.

Table 5

	All Production		Above Median Production		Below Median Production	
	Oil Prod	Wealth Est	Oil Prod	Wealth Est	Oil Prod	Wealth Est
Mean	475,484	$1,405,052	933,152	$2,748,020	23,032	$88,945
SD	1,417,713	$3,970,957	1,904,624	$5,319,152	20,419	$80,212
p50	76,335	$247,524	295,694	$1,059,840	17,048	$62,966
N	2,094		1,041		1,053	
% of Treated	93.19%		46.33%		46.86%	

Notes: This table reports the mean, median, and standard deviation of oil production among children from treated sections with positive production (a small number of treated children have zero production reported in USGS). Production is imputed by allocating the cumulative 1968 oil production in every two-mile-by-two-mile cell equally among each of its one-mile-by-one-mile cells with 5 or more USGS wells by 1968. We assign each one-mile-by-one-mile cell to the closest section and divide the production amount by four (the typical number of homesteaders in a section). Column 1 includes all treated children with positive production. Columns 2 and 3 divide children into those above and below the median. Production volumes are converted into wealth by taking 12.5%—the typical royalty rate—of the imputed oil production multiplied by the price of oil in the year of discovery. For comparison, median annual earnings in 1940 are $16,880 among of our sample of untreated children (all dollar amounts are expressed in 2023 dollars).

Table 6

Effects of parental oil wealth on sons' wealth in 1940

	Owned home		Home value > $5000		Nonwage income > $50	
	(1)	(2)	(3)	(4)	(5)	(6)
Odds ratios						
Parent oil	0.162**	0.180*	0.255*	0.463***	0.010	0.218*
	(0.077)	(0.097)	(0.148)	(0.176)	(0.086)	(0.114)
Parent oil x parent is under 70 in 1940	-0.239	-0.309	0.194	0.056	-0.008	-0.612*
	(0.188)	(0.242)	(0.429)	(0.573)	(0.216)	(0.317)
Above-median wealth	No	Yes	No	Yes	No	Yes
Township FE	Yes		Yes		Yes	
Control for birth cohort	Yes		Yes		Yes	
Control for baseline interactions	Yes		Yes		Yes	
Observations	7,041	6,368	6,669	5,977	7,035	6,366
Average marginal effects						
Parent is 70 or older in 1940	0.040**	0.044*	0.014	0.028**	0.003	0.054*
	(0.019)	(0.024)	(0.008)	(0.013)	(0.021)	(0.028)
Parent is under 70 in 1940	-0.019	-0.032	0.024	0.030	0.001	-0.092
	(0.043)	(0.056)	(0.025)	(0.037)	(0.051)	(0.068)
Above-median wealth	No	Yes	No	Yes	No	Yes
Mean outcome with no oil, parent is 70 or older in 1940	0.547		0.057		0.487	

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

Notes: The top panel displays intent-to-treat odds ratios from logistic regression models estimated on the agreement sample, with "parent oil" as an indicator for having a parent who lived in a section by 1910 with an oil discovery by 1940. Standard errors (reported in parentheses) are clustered as the section level, which is one square mile and contains about four households. Each model includes fixed effects for townships, which each contain 36 sections. Additional control variables include age and indicators for race. Models with "above-median wealth" restrict the intention-to-treat group to children whose parent lived in a section with a value of oil production above the median. We restrict the sample to children from townships with oil discoveries who were at least 20 years old in 1940. The bottom panel displays predicted marginal effects (with delta-method standard errors in parentheses) of having a parent who lived in a section by 1910 with an oil discovery by 1940, relative to individuals within the township but without oil.

Table 7

Effects of parental oil wealth on daughters' wealth in 1940

	Owned home		Home value > $5000		Nonwage income > $50	
	(1)	(2)	(3)	(4)	(5)	(6)
Odds ratios						
Parent oil	0.082	0.239*	0.333**	0.613***	0.248*	0.450***
	(0.097)	(0.129)	(0.169)	(0.198)	(0.133)	(0.166)
Parent oil x parent is under 70 in 1940	-0.148	-0.051	-0.489	-0.604	-0.458	-0.949
	(0.240)	(0.326)	(0.532)	(0.679)	(0.455)	(0.784)
Above-median wealth	No	Yes	No	Yes	No	Yes
Township FE	Yes		Yes		Yes	
Control for birth cohort	Yes		Yes		Yes	
Control for baseline interactions	Yes		Yes		Yes	
Observations	4,771	4,332	4,588	4,067	4,718	4,277
Average marginal effects						
Parent is 70 or older in 1940	0.020	0.057*	0.023*	0.050**	0.035*	0.068**
	(0.023)	(0.030)	(0.013)	(0.020)	(0.020)	(0.028)
Parent is under 70 in 1940	-0.016	0.045	-0.007	0.000	-0.030	-0.066
	(0.056)	(0.074)	(0.021)	(0.032)	(0.059)	(0.086)
Above-median wealth	No	Yes	No	Yes	No	Yes
Mean outcome with no oil, parent is 70 or older in 1940	0.592		0.075		0.190	

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

Notes: The top panel displays intent-to-treat odds ratios from logistic regression models estimated on the agreement sample, with "parent oil" as an indicator for having a parent who lived in a section by 1910 with an oil discovery by 1940. Standard errors (reported in parentheses) are clustered as the section level, which is one square mile and contains about four households. Each model includes fixed effects for townships, which each contain 36 sections. Additional control variables include age and indicators for race. Models with "above-median wealth" restrict the intention-to-treat group to children whose parent lived in a section with a value of oil production above the median. We restrict the sample to children from townships with oil discoveries who were at least 20 years old in 1940. The bottom panel displays predicted marginal effects (with delta-method standard errors in parentheses) of having a parent who lived in a section by 1910 with an oil discovery by 1940, relative to individuals within the township but without oil.

Table 8

Effects of parental oil wealth on sons' education, labor supply, and migration in 1940

	OLS regressions								Logistic regressions (odds ratios)					
	Years of education		Weekly hours worked		Annual wage income		Spouse's education		In labor force		White collar job		In a city	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Parent oil	1.025	0.967	5.874	1.408	156.258	222.573	-0.265	-0.486	1.077	0.069	-0.161	0.936	0.924*	1.978**
	(0.635)	(1.186)	(5.223)	(6.353)	(181.024)	(289.068)	(0.665)	(1.051)	(1.029)	(1.199)	(0.529)	(0.627)	(0.508)	(0.860)
Parent oil x age at township discovery	-0.055	-0.045	-0.537	-0.194	-4.441	-10.774	0.028	0.067	-0.091	-0.07	0.027	-0.045	-0.038	-0.095
	(0.048)	(0.091)	(0.382)	(0.492)	(13.776)	(23.148)	(0.049)	(0.081)	(0.077)	(0.105)	(0.037)	(0.052)	(0.037)	(0.065)
Parent oil x township discovery after age 18	-0.456	-0.171	-2.592	4.362	-45.406	-290.366	0.602	1.216	-1.459	-0.925	0.089	-1.193*	-1.226**	-2.131**
	(0.766)	(1.329)	(5.853)	(7.381)	(213.386)	(325.852)	(0.751)	(1.168)	(1.113)	(1.330)	(0.603)	(0.723)	(0.566)	(0.909)
Parent oil x age at township discovery	0.04	0.025	0.43	0.022	0.544	11.485	-0.039	-0.083	0.097	0.087	-0.025	0.057	0.048	0.102
x township discovery after age 18	(0.049)	(0.093)	(0.389)	(0.508)	(14.123)	(23.648)	(0.050)	(0.082)	(0.078)	(0.106)	(0.038)	(0.054)	(0.038)	(0.066)
Above-median wealth	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Township FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Control for birth cohort	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Control for baseline interactions	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	6,771	6,130	7,041	6,368	6,336	5,727	5,424	4,901	6,966	6,278	7,016	6,337	7,023	6,350

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

Notes: Intent-to-treat coefficients (from OLS models) or odds ratios (from logistic regression models) estimated on the agreement sample, with "parent oil" as an indicator for having a parent who lived in a section by 1910 with an oil discovery by 1940. Standard errors (reported in parentheses) are clustered as the section level, which is one square mile and contains about four households. Each model includes fixed effects for townships, which each contain 36 sections. Age at township discovery is defined using the date when most oil discoveries occurred within the township, and is independent of the presence of oil in the section. Additional control variables include age and indicators for race. Models with "above-median wealth" restrict the intention-to-treat group to children whose parent lived in a section with a value of oil production above the median. We restrict the sample to children from townships with oil discoveries who were at least 20 years old in 1940.

Table 9

Effects of parental oil wealth on daughters' education, labor supply, and migration in 1940

	OLS regressions								Logistic regressions (odds ratios)					
	Years of education		Weekly hours worked		Annual wage income		Spouse's education		In labor force		White collar job		In a city	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Parent oil	0.664	-0.149	1.836	4.846	131.824	124.665	1.748*	1.551	0.529	0.438	1.122*	1.067	-0.572	-0.038
	(0.672)	(1.058)	(5.386)	(7.012)	(103.635)	(139.394)	(1.019)	(1.577)	(0.586)	(0.804)	(0.640)	(0.769)	(0.566)	(0.879)
Parent oil x age at township discovery	-0.024	0.049	-0.193	-0.248	-9.718	-7.996	-0.085	-0.071	-0.05	-0.02	-0.071	-0.042	0.041	0.008
	(0.045)	(0.071)	(0.388)	(0.522)	(7.388)	(9.933)	(0.074)	(0.113)	(0.047)	(0.062)	(0.049)	(0.057)	(0.040)	(0.065)
Parent oil x township discovery after age 18	-0.094	1.362	-1.791	-5.292	-142.922	-67.765	-0.698	-0.374	-0.245	-0.032	-0.924	-0.673	0.812	0.274
	(0.805)	(1.179)	(5.762)	(7.601)	(115.935)	(157.992)	(1.200)	(1.756)	(0.707)	(0.975)	(0.807)	(1.007)	(0.616)	(0.940)
Parent oil x age at township discovery	0.005	-0.085	0.175	0.238	9.624	5.915	0.056	0.037	0.046	0.013	0.066	0.04	-0.046	-0.016
x township discovery after age 18	(0.047)	(0.073)	(0.391)	(0.531)	(7.528)	(10.154)	(0.077)	(0.116)	(0.048)	(0.064)	(0.051)	(0.060)	(0.041)	(0.065)
Above-median wealth	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Township FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Control for birth cohort	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Control for baseline interactions	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	4,637	4,214	4,779	4,339	3,943	3,578	3,476	3,156	4,718	4,279	4,637	4,221	4,771	4,328

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

Notes: Intent-to-treat coefficients (from OLS models) or odds ratios (from logistic regression models) estimated on the agreement sample, with "parent oil" as an indicator for having a parent who lived in a section by 1910 with an oil discovery by 1940. Standard errors (reported in parentheses) are clustered as the section level, which is one square mile and contains about four households. Each model includes fixed effects for townships, which each contain 36 sections. Age at township discovery is defined using the date when most oil discoveries occurred within the township, and is independent of the presence of oil in the section. Additional control variables include age and indicators for race. Models with "above-median wealth" restrict the intention-to-treat group to children whose parent lived in a section with a value of oil production above the median. We restrict the sample to children from townships with oil discoveries who were at least 20 years old in 1940.

Table 10

Predicted effects of parental oil wealth on sons' education, labor supply, and migration in 1940

	Years of education		Weekly hours worked		Annual wage income		Spouse's education		In labor force		White collar job		In a city	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Township discovery at age 0	1.025	0.967	5.874	1.408	156.258	222.573	-0.265	-0.486	0.027	0.003	-0.015	0.128	0.217*	0.457***
	(0.635)	(1.186)	(5.223)	(6.353)	(181.024)	(289.068)	(0.665)	(1.051)	(0.022)	(0.047)	(0.049)	(0.103)	(0.121)	(0.163)
Township discovery at age 5	0.747*	0.743	3.192	0.439	134.054	168.703	-0.123	-0.149	0.016	-0.011	-0.003	0.110	0.176**	0.359***
	(0.441)	(0.806)	(3.505)	(4.251)	(120.859)	(185.274)	(0.449)	(0.704)	(0.015)	(0.032)	(0.046)	(0.074)	(0.083)	(0.120)
Township discovery at age 10	0.470	0.518	0.509	-0.530	111.850	114.834	0.019	0.187	0.005	-0.026	0.018	0.084	0.132**	0.251***
	(0.313)	(0.553)	(2.138)	(2.808)	(77.025)	(109.017)	(0.289)	(0.459)	(0.011)	(0.020)	(0.038)	(0.062)	(0.051)	(0.071)
Township discovery at age 15	0.193	0.294	-2.174	-1.500	89.646	60.964	0.161	0.524	-0.008	-0.040	0.049	0.049	0.087*	0.135*
	(0.338)	(0.610)	(2.036)	(3.130)	(82.146)	(127.417)	(0.297)	(0.502)	(0.012)	(0.036)	(0.044)	(0.084)	(0.049)	(0.071)
Township discovery at age 20	0.256	0.394	1.159	2.339	32.907	-53.566	0.136	0.413	-0.004	-0.008	-0.005	-0.006	-0.023	-0.003
	(0.222)	(0.291)	(1.440)	(1.907)	(63.381)	(79.291)	(0.220)	(0.261)	(0.004)	(0.006)	(0.032)	(0.038)	(0.033)	(0.042)
Township discovery at age 25	0.178	0.294	0.628	1.481	13.421	-50.010	0.085	0.333	-0.005	-0.010	-0.002	0.006	-0.010	0.006
	(0.186)	(0.243)	(1.199)	(1.576)	(54.062)	(67.192)	(0.186)	(0.218)	(0.006)	(0.008)	(0.025)	(0.031)	(0.028)	(0.035)
Township discovery at age 30	0.099	0.193	0.097	0.624	-6.066	-46.453	0.035	0.254	-0.008	-0.013	0.000	0.017	0.003	0.014
	(0.158)	(0.206)	(1.025)	(1.339)	(46.157)	(57.367)	(0.158)	(0.185)	(0.008)	(0.010)	(0.020)	(0.025)	(0.023)	(0.029)
Above-median wealth	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Mean outcome with no oil, township discovery by age 30	9.794		37.143		836.634		10.300		0.950		0.240		0.402	

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

Notes: Predicted marginal effects of having a parent who lived in a section by 1910 with an oil discovery by 1940, relative to individuals within the township but without oil. Marginal effects are reported by the age when most oil discoveries occurred within the township. Standard errors (reported in parentheses) are generated using the delta method. Models with "above-median wealth" restrict the intention-to-treat group to children whose parent lived in a section with a value of oil production above the median.

Table 11

Predicted effects of parental oil wealth on daughters' education, labor supply, and migration in 1940

	Years of education		Weekly hours worked		Annual wage income		Spouse's education		In labor force		White collar job		In a city	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Township discovery at age 0	0.664	-0.149	1.836	4.846	131.824	124.665	1.748*	1.551	0.131	0.109	0.257*	0.246	-0.142	-0.009
	(0.672)	(1.058)	(5.386)	(7.012)	(103.635)	(139.394)	(1.019)	(1.577)	(0.143)	(0.197)	(0.144)	(0.181)	(0.138)	(0.218)
Township discovery at age 5	0.545	0.096	0.871	3.604	83.234	84.684	1.324*	1.199	0.066	0.083	0.155*	0.180	-0.092	0.001
	(0.483)	(0.750)	(3.604)	(4.737)	(69.689)	(94.264)	(0.687)	(1.074)	(0.092)	(0.131)	(0.089)	(0.120)	(0.094)	(0.145)
Township discovery at age 10	0.425	0.342	-0.094	2.362	34.643	44.704	0.901**	0.846	0.005	0.054	0.068	0.117	-0.041	0.011
	(0.343)	(0.510)	(2.118)	(3.055)	(41.256)	(57.213)	(0.427)	(0.675)	(0.050)	(0.075)	(0.049)	(0.077)	(0.057)	(0.082)
Township discovery at age 15	0.305	0.587	-1.059	1.120	-13.947	4.723	0.477	0.494	-0.039	0.028	0.008	0.065	0.009	0.021
	(0.319)	(0.461)	(1.868)	(3.139)	(35.731)	(50.929)	(0.409)	(0.631)	(0.046)	(0.070)	(0.040)	(0.067)	(0.051)	(0.074)
Township discovery at age 20	0.197	0.493	-0.308	-0.647	-12.978	15.286	0.474	0.498	0.036	0.054	0.015	0.054	0.032	0.019
	(0.257)	(0.323)	(1.201)	(1.641)	(27.128)	(37.539)	(0.292)	(0.389)	(0.039)	(0.055)	(0.037)	(0.056)	(0.038)	(0.053)
Township discovery at age 25	0.104	0.313	-0.397	-0.697	-13.447	4.882	0.330	0.328	0.028	0.043	0.010	0.043	0.025	0.009
	(0.213)	(0.260)	(0.993)	(1.325)	(22.835)	(30.490)	(0.236)	(0.312)	(0.029)	(0.039)	(0.024)	(0.036)	(0.031)	(0.043)
Township discovery at age 30	0.011	0.133	-0.485	-0.747	-13.917	-5.521	0.186	0.158	0.022	0.033	0.006	0.034	0.018	-0.001
	(0.180)	(0.216)	(0.827)	(1.074)	(19.312)	(24.310)	(0.196)	(0.253)	(0.022)	(0.028)	(0.017)	(0.024)	(0.026)	(0.035)
Above-median wealth	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Mean outcome with no oil, township discovery by age 30	10.545		8.659		152.264		9.553		0.214		0.136		0.440	

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

Notes: Predicted marginal effects of having a parent who lived in a section by 1910 with an oil discovery by 1940, relative to individuals within the township but without oil. Marginal effects are reported by the age when most oil discoveries occurred within the township. Standard errors (reported in parentheses) are generated using the delta method. Models with "above-median wealth" restrict the intention-to-treat group to children whose parent lived in a section with a value of oil production above the median.

Appendix A Cushing Oil Field

We have included a map of the three data sources compared to a 1918 map of the Cushing oil field because of the data's abilities to capture its unique shape. Although the oil field is not in Oklahoma Territory and so is not in our final children's sample, it represents a large early Oklahoma oil field and shows the pros and cons of the different oil data sources and our method of treatment in an uniquely shaped oil field. As with the Oklahoma City field, each dataset traces the contours of the field reasonably well, with minor disagreements around the boundaries of the field. Disagreements are most frequent around the edges of apparent oil fields, which may reflect exploratory wells that came up dry and were thus less likely to be recorded in some data sources. Figure 3 color-codes each section of Oklahoma Territory according to how many data sources indicate the presence of oil in that section. Our labels closely align with the historical maps in Figures 9 and 9d, indicating that we can minimize both false positives and negatives by combining information from OGS, USGS, and WellDatabase.

Appendix B Bounding and bias-corrected estimates

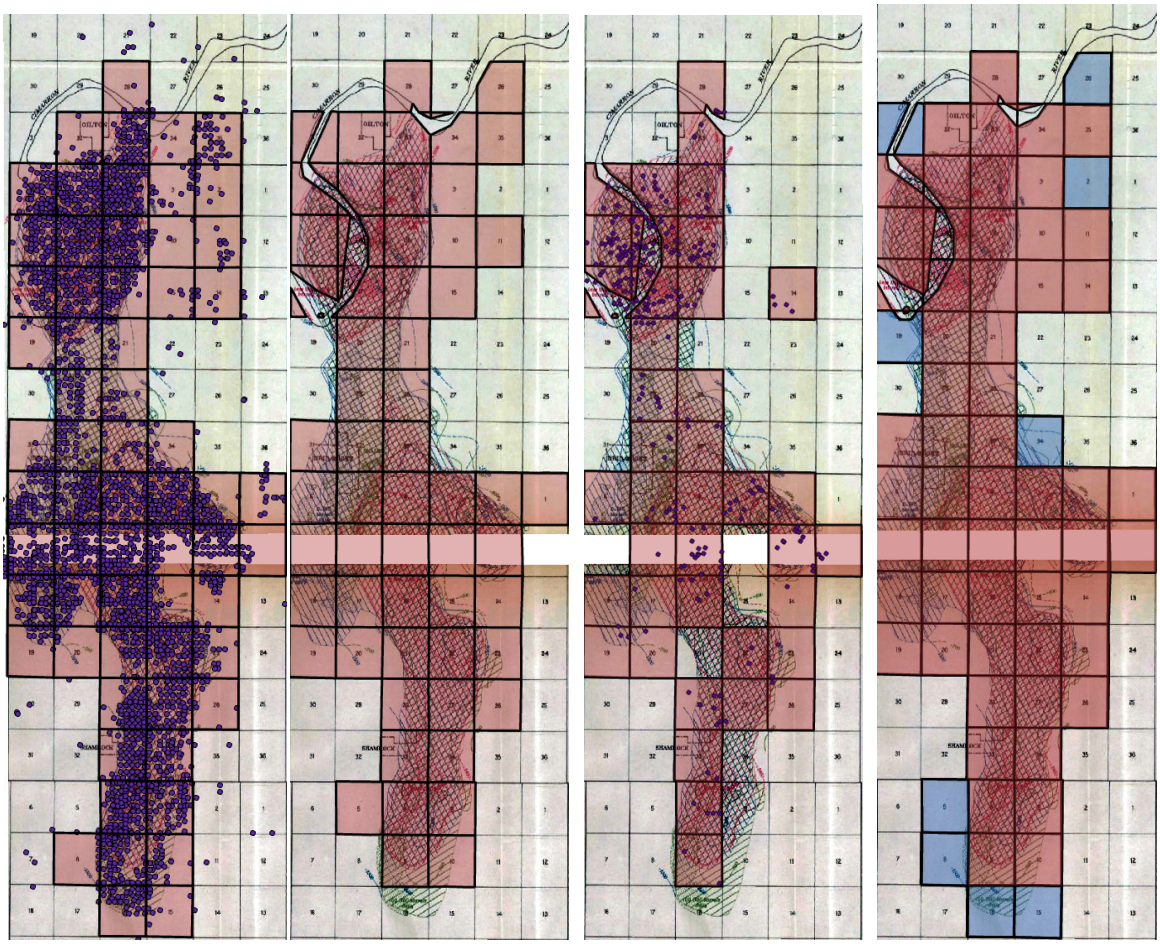
In this appendix, we report estimates using the bounding and bias-correction approaches suggested by Ferrara et al. (2024). They show that measurement error in a binary right-hand side variable will bias OLS estimates toward zero while the instrumental variables estimator (using one noisy measure as an instrument for the other) is biased away from zero. Thus, OLS and IV estimates provide bounds on the true effect. In Table 12, 13, and 14 we report both OLS and IV estimates where the independent variable is a dummy for whether the section has five or more wells in the OGS data Z^{OGS} and the instruments are dummies based on the USGS and WellDatabase data Z^{USGS} and Z^{WD}. We estimate the following OLS regression

$$y_{ijt} = \delta_0 + \delta^{OLS} Z_j^{OGS} + X_i \cdot \beta + a_t + \epsilon_{ijt}$$

which provides a lower bound on the treatment effect while its IV counterpart δ^{IV} (using Z^{USGS} and Z^{WD} as instruments for Z^{OGS}) provides an upper bound. Ferrara et al. (2024) also suggest a bias-corrected estimator $\delta^{BC} = \sqrt{\delta^{OLS} \times \delta^{IV}}$, however this estimator has the unappealing property of being undefined when the OLS and IV estimates happen to be of opposite sign. Nevertheless, Table 12, 13, and 14 also reports bias-corrected estimates when they are defined.

Figure 9: Cushing Oil Field Map by 1918

(a) OK Geol. Survey (b) U.S. Geol. 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 only one dataset shows oil is present. Red squares represent sections where two or more datasets agree oil is present.

Table 12: Intent-to-treat effects of oil discovery on children's wealth in 1940

	Home				Home Value > $5000				Nonwage Income > $50			
	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}
Oil in section (Oklahoma Geological Survey)	0.008		0.027*	0.052**	0.014*		0.022***	0.037***	0.013		0.018	0.022
	(0.015)		(0.016)	(0.023)	(0.008)		(0.008)	(0.014)	(0.010)		(0.011)	(0.017)
Oil in section (2-3 sources)		-0.013				-0.017				-0.016		
		(0.016)				(0.016)				(0.016)		
Mean outcome, no oil		0.531				0.049				0.150		
Observations	90,325	90,309	-	90,309	90,325	90,309	-	90,309	90,325	90,309	-	90,309

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

Notes: OLS, AS, and 2SLS models with township fixed effects and standard errors (in parentheses) clustered by section. We instrument for having 5+ wells in the Oklahoma Geological Survey using 1+ wells in the WellDatabase and the 5+ wells U.S. Geological Survey. 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. The agreement sample has the some conditions but drops sections where only one dataset reports oil in that section. Oil in section is defined using drilled wells from three sources for which each has a different cutoff (as described above). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 13: Intent-to-treat effects of oil discovery on children's income and labor supply in 1940

	Wage income				Hours worked				In labor force			
	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}
Oil in section (Oklahoma Geological Survey)	-11.150		-14.110	-38.251	-0.419		-0.464	-1.323	-0.004		-0.002	-0.014
	(25.555)		(25.899)	(37.404)	(0.599)		(0.654)	(0.974)	(0.009)		(0.010)	(0.015)
Oil in section (2-3 sources)		0.469				0.008				-0.012		
		(0.592)				(0.105)				(0.014)		
Mean outcome, no oil		415.716				23.840				0.600		
Observations	81,589	90,309	-	81,571	90,325	90,309	-	90,309	90,325	90,309	-	90,309

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

Notes: OLS, AS, and 2SLS models with township fixed effects and standard errors (in parentheses) clustered by section. We instrument for having 5+ wells in the Oklahoma Geological Survey using 1+ wells in the WellDatabase and the 5+ wells U.S. Geological Survey. 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. The agreement sample has the some conditions but drops sections where only one dataset reports oil in that section. Oil in section is defined using drilled wells from three sources for which each has a different cutoff (as described above). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.

Table 14: Intent-to-treat effects of oil discovery on children's education and migration in 1940

	Years of Education				Oklahoma				Urban			
	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}	δ^{OLS}	δ^{AS}	δ^{BC}	δ^{IV}
Oil in section (Oklahoma Geological Survey)	0.058		0.020	-0.075	0.011		0.026	0.046	0.021		0.019	0.024
	(0.124)		(0.129)	(0.189)	(0.019)		(0.020)	(0.030)	(0.016)		(0.017)	(0.026)
Oil in section (2-3 sources)		-0.034				-0.015				-0.020		
		(0.043)				(0.016)				(0.017)		
Mean outcome, no oil		9.151				0.484				0.384		
Observations	87,743	90,309	-	87,727	90,325	90,309	-	90,309	90,325	90,309	-	90,309

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

Notes: OLS, AS, and 2SLS models with township fixed effects and standard errors (in parentheses) clustered by section. We instrument for having 5+ wells in the Oklahoma Geological Survey using 1+ wells in the WellDatabase and the 5+ wells U.S. Geological Survey. 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. The agreement sample has the some conditions but drops sections where only one dataset reports oil in that section. Oil in section is defined using drilled wells from three sources for which each has a different cutoff (as described above). A section is one square mile (with up to 4 original homesteads) and a township comprises 36 sections.