Motivation and Contribution

- Recent advances in shale development have produced both positive and negative outcomes for local communities, with higher employment and income known to be the most significant effects.
- Despite the stated importance of the distribution of economic gains among local populations in previous literature, adequate research on the shale boom’s impact on inequality and affordability does not exist.
- I employ the difference-in-difference (DiD) method to study the unintended social consequences of the hydraulic fracturing boom in Oklahoma, the second-largest producer of oil and gas in the country, over the period of 2004-2017.

Methodology

- The difference-in-difference (DiD) method: identifying a specific intervention or treatment and then comparing the difference in outcomes after and before the intervention for groups affected by it to the same difference for unaffected groups.
- I define the treatment group to be counties that experienced shale extraction and the treatment to be the shale development boom that happened in 2008.
- To avoid endogeneity, shale counties were selected based on their geographic location, as it is common in the literature.
- To avoid spillover effect, I construct and report results for a second sample that excludes non-shale counties that share a border with a shale county.
- Moreover, I apply same estimation to a third data set with all the counties except the two metropolitan counties, Oklahoma county and Tulsa county, to avoid their excessive influence on the regression results.
- The DiD estimator is based on a strong identifying assumption: the availability of a treatment and a control group that would have had a similar trend without the treatment. To address this issue, I include a set of covariates; these covariates are used to describe how the average effect of the treatment varies with changes in observed characteristics. I also follow the literature (e.g., Linden and Rockoff, 2008; Caselli and Michaels, 2013) to examine preexisting differences in counties’ characteristics using a cross-sectional estimator.
- Given the similarity of pre-treatment trends in shale and non-shale counties, I then use a linear DiD model to estimate shale boom effect on local communities:

\[ \text{ln}(\text{Outcome}_{ct}) = \beta_0 + \beta_1 \text{Shale}_{c} + \beta_2 \text{Post}_{2008} + \beta_3 \text{Shale}_{c} \times \text{Post}_{2008} + \mu_c + \tau_t + \epsilon_{ct} \]

The dependent variable, ln(Outcome_{ct}), represents outcomes of interest: housing price index, Gini coefficient, and housing affordability index. \( \beta_1 \) is the coefficient of interest which measures the average shale development effect on shale counties by differencing the changes in outcomes in shale counties after 2008 with non-shale counties.
- I included income per capita, housing density and population density using the Census 2010 data as a set of control variables that allows counties with different characteristics to have different outcome. To make the specification even more robust, I follow the literature to include control variables for neighboring counties as well:

\[ \text{ln}(\text{Outcome}_{ct}) = \beta_0 + \beta_1 \text{Shale}_{c} + \beta_2 \text{Post}_{2008} + \beta_3 \text{Shale}_{c} \times \text{Post}_{2008} + \beta_4 X_c + \beta_5 C_c + \mu_c + \tau_t + \epsilon_{ct} \]

where \( X_c \) is a set of control variables capturing counties’ observable characteristics and \( C_c \) is neighboring counties’ observable characteristics.

References


Methodology

- The difference-in-difference (DiD) method: identifying a specific intervention or treatment and then comparing the difference in outcomes after and before the intervention for groups affected by it to the same difference for unaffected groups.
- I define the treatment group to be counties that experienced shale extraction and the treatment to be the shale development boom that happened in 2008.
- To avoid endogeneity, shale counties were selected based on their geographic location, as it is common in the literature.
- To avoid spillover effect, I construct and report results for a second sample that excludes non-shale counties that share a border with a shale county.
- Moreover, I apply same estimation to a third data set with all the counties except the two metropolitan counties, Oklahoma county and Tulsa county, to avoid their excessive influence on the regression results.
- The DiD estimator is based on a strong identifying assumption: the availability of a treatment and a control group that would have had a similar trend without the treatment. To address this issue, I include a set of covariates; these covariates are used to describe how the average effect of the treatment varies with changes in observed characteristics. I also follow the literature (e.g., Linden and Rockoff, 2008; Caselli and Michaels, 2013) to examine preexisting differences in counties’ characteristics using a cross-sectional estimator.
- Given the similarity of pre-treatment trends in shale and non-shale counties, I then use a linear DiD model to estimate shale boom effect on local communities:

\[ \text{ln}(\text{Outcome}_{ct}) = \beta_0 + \beta_1 \text{Shale}_{c} + \beta_2 \text{Post}_{2008} + \beta_3 \text{Shale}_{c} \times \text{Post}_{2008} + \mu_c + \tau_t + \epsilon_{ct} \]

The dependent variable, ln(Outcome_{ct}), represents outcomes of interest: housing price index, Gini coefficient, and housing affordability index. \( \beta_1 \) is the coefficient of interest which measures the average shale development effect on shale counties by differencing the changes in outcomes in shale counties after 2008 with non-shale counties.
- I included income per capita, housing density and population density using the Census 2010 data as a set of control variables that allows counties with different characteristics to have different outcome. To make the specification even more robust, I follow the literature to include control variables for neighboring counties as well:

\[ \text{ln}(\text{Outcome}_{ct}) = \beta_0 + \beta_1 \text{Shale}_{c} + \beta_2 \text{Post}_{2008} + \beta_3 \text{Shale}_{c} \times \text{Post}_{2008} + \beta_4 X_c + \beta_5 C_c + \mu_c + \tau_t + \epsilon_{ct} \]

where \( X_c \) is a set of control variables capturing counties’ observable characteristics and \( C_c \) is neighboring counties’ observable characteristics.

Results

<table>
<thead>
<tr>
<th></th>
<th>Housing Price Index</th>
<th>Gini Coefficient</th>
<th>Housing Affordability Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Post 2008</td>
<td>0.0028</td>
<td>0.0070</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Shale County</td>
<td>0.1099***</td>
<td>0.0243</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Shale x Post 2008</td>
<td>0.0514**</td>
<td>0.0534**</td>
<td>0.0480**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.179**</td>
<td>4.120***</td>
<td>3.411**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>County Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year x Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>County Set of Control Variables: Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Neighboring County Set of Control Variables: Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Number of Observations: 684 684 684 686 686 630 684 684 628
Adj. R^2: 0.034 0.033 0.024 0.094 0.075 0.076 0.082 0.083 0.097

Standard errors are in parentheses:** *p < 0.01, **p < 0.05, ***p < 0.005

Data

- For housing values: county-level annual housing price index (HPI) from the Federal Housing Finance Agency (FHFA)
- For inequality: Internal Revenue Service (IRS) data for the Gini index
- For affordability: the median household income from the IRS data and the real housing price index from FHFA for the frequently used housing affordability index, the income-to-housing price ratio
- For counties’ characteristics: Census 2010 data (income per capita, population density, and housing density)

Conclusion

- The results suggest that the shale boom was associated with appreciation in housing values for 5.5% and a decrease of 9% in affordability in the state of Oklahoma for shale counties compared to non-shale counties.
- Although previous literature provides evidence for higher employment and income in shale counties due to the boom, the estimation fails to find any statistically significant effect on inequality.
- The results are consistent across three different samples, with or without covariates for shale counties or their neighbors.