The Dark Side of the Shale Boom– Increased Crime among Economically-small, Relatively Rural American States: Generalized Synthetic Control Approach[☆]

Shishir Shakya
College of Business and Economics and Regional Research Institute,
West Virginia University, Morgantown, WV 26505
E-mail: shishir.shakya@mail.wvu.edu

Abstract: Literatures relate sudden expansion of tight oil and shale gas in the US with higher employment and income. However, dis-amenities like crime rates attract little attention. This paper exploits the US shale boom as a natural experiment and compares shale-infeasible states with economically-small, relatively rural shale boom states to investigate impacts on crimes and victimization costs. I utilize the double-selection post-LASSO method for proper selection on observable confounders and the Generalize Synthetic Control (GSC) approach to control the unobserved time-varying confounding effect. Unlike, synthetic control, GSC can estimate confidence intervals for counterfactual and incorporate multiple treatment units. This method is applicable for causal inference in the regional policy setting when treatment occurs in several states at different periods. My results show a significant rise in violent crimes among treatment states. I provide falsification tests for pretreatment trend and sample selection along with the robustness checks. My results support a causal interpretation. I estimate 15.68 million (2008 dollar) worth of the annual victimization cost of the fracking boom for the treatment states.

Keywords: crime, fracking boom, victimization cost

JEL Codes: Q33 · R11 · K42

[☆] I am grateful to Prof. Randall Jackson, Caigan Mckenzie, Prof. Brad R. Humphreys for helpful comments and guidance in this work.

1. Introduction

The recent technological development of hydraulic fracturing and horizontal drilling (also known as fracking) have become one of the most economically important innovations in the energy industry in the past 30 years. Fracking technology made it feasible to extract previously inaccessible natural gas and tight oil from the shale formations. As the U.S. holds 70 years of natural gas reserves (EIA, 2013), it is forecasted to overtake Saudi Arabia and Russia as the world's leading energy producer by 2019 (Meredith, 2018). Fracking boom revitalizes many states and local economies in the U.S. with increased employment and income. Several studies examine the income and employment aspect of the shale boom; while local dis-amenities like crime rates attract little attention.

Komarek (2018) highlights several adverse effects of the sudden shale boom in the US as reported in mainstream media outlets. For example, the National Geographic points to increases in binge drinking, prostitution and violence in North Dakota (Dobb, 2013), The Washington Post notes a rise in drug-related crimes in Washington (Horwitz, 2014), The New York Times reports increased violence against women (Eligon, 2013), fatal accidents and sexually transmitted infections (Food and Water Watch, 2013). This paper exploits shale boom as the natural experiment to examine crimes (Part-I and Part-II offenses as classified by Uniform Crime Report) in the relatively rural states like West Virginia, North Dakota, and Arkansas. These states are economically-small, relatively more rural, and are equally affected by shale oil and gas boom.

Becker's (1968) seminal work on an economic approach to crime and punishment postulates that criminals are rational economic agents and they engage in criminal activities when benefits outweigh the costs of committing a crime. The cost of committing a crime in Becker's model includes the probability of apprehension and imprisonment, and the opportunity cost of legitimate employment, among other expenses (Komarek, 2018). The resource boom increases wage and employment opportunities, which increases the opportunity cost of apprehension, thus, it should reduce criminal activities. Grogger (1998) provides a theoretical prediction that resource boom leads to increase in wages thus, reduces crime. Gould, Weinberg, & Mustard (2002) show that resource boom leads to increased employment and wages of low-skilled individuals, therefore, it reduces crime. Furthermore, Komarek (2016) shows evidence of growing jobs and wages in non-extraction sectors like construction, transportation, retail trade, and accommodations sector due to fracking boom in Marcellus region. Similarly, Allcott & Keniston (2018) find increase employment and wages in manufacturing. Other than wage effect, Grinols & Mustard (2006) state that the resource boom leads to renovations like more street lights and the presence of residents. Therefore, such developments should indirectly reduce crime rates.

However, Grinols & Mustard (2006) also provide several competing arguments that resource boom can increase local crime via indirect channels of development, the increased payoff from crime, visitor criminality, and change in population composition. Regarding the fracking boom, mostly, the non-residents take the extraction-related new jobs, most of which are temporary (Ruddell et al., 2014). The non-residents who are likely to be drive-out or fly-in and fly-out employees have a higher opportunity cost for committing a crime (Komarek, 2018). The sudden upsurge in employment and labor demand also attracts heterogeneous unemployed and underemployed individuals. James & Smith (2017) point out that "among a host of factors that determine labor market success rates, a history of drug use and a recorded criminal conviction significantly reduce an individual's ability to find employment.".

This study investigates the effect of the fracking boom in the local crime within and across the states of West Virginia, North Dakota, and Arkansas. These states are economically-small and relatively more rural American states. This study contributes to the literature to measure the effect of a natural resource extraction boom on criminal activity and complements James & Smith (2017) and Komarek (2018). Unlike, James & Smith (2017) who quantify the cost of shale boom-related crimes for the USA in aggregate level, this paper provides cost of shale boom-related crimes for West Virginia, North Dakota, and Arkansas. Contrary to Komarek (2018) who uses difference-in-differences estimation as identification strategy with a panel of 2004-2012, this paper merges state-level data from several sources to develop high-dimensional panel data from 1999-2015 and is more rigorous on identification and estimation strategy.

My identification strategy exploits a natural experiment of the fracking boom. I use Belloni et al. (2013) method of double-selection post-LASSO for proper selection on observables confounders. Then, I implement Xu (2017) method of Generalize Synthetic Control (GSC) to absorb the unobserved time-varying confounding effect. Unlike, a synthetic control method (Abadie & Gardeazabal, 2003; Miller et al., 2010), GSC can estimate confidence intervals for counterfactual and incorporate multiple treatment units. GSC also embeds Athey et al. (2017) matrix completion method for causal panel data to estimate counterfactual. The double-selection post-LASSO and matrix completion are machine learning approaches redesigned for causal inference.

The next section presents the detailed identification strategy. Section 3 extends the empirical approach, including discussion of the data, double-selection post-Lasso, and GSC. Section 4 displays the results and discuss the mechanisms followed by falsification tests, robustness checks and estimates of victimization cost of the fracking boom. Section 5 concludes the study.

2. Identification Strategy

In this section, I explain the sample selection then I illustrate my identification[1] strategy. Endogeneity biases the estimation and threatens the causal estimates in observational studies. Endogeneity may arise via reverse causality, omitted variables and unobservable factors. I explain my strategy for coping with each source of endogeneity.

Sample selection: treatment and control group

In this study, the treatment group comprises West Virginia (51.28%), North Dakota (40%) and Arkansas (43.84%). The parenthesis encloses the percentage of the population residing in the rural area as reported in the 2010 Census Urban and Rural Classification and Urban Area Criteria. These states experience shale oil and gas boom and are economically-small and relatively rural. In this study, the control group is infeasible for the shale production namely the state of Arizona, Georgia, Idaho, Iowa, Minnesota, Missouri, Oregon, South Carolina, South Dakota, Washington, and Wisconsin (Hirji & Song, 2016).

Endogeneity

The main research question is whether the treatment variable D (resource boom due to fracking) affects the outcome variable Y (local crime) or not. Therefore, the causal direction flows from D to Y given as $D \rightarrow Y$. It is highly unlikely that crime affects the fracking boom, hence the *endogeneity from reverse causality* $D \leftarrow Y$ is unlikely.

After the fracking innovation is feasible, the fracking boom occurs with political and business motives which possibly depends upon the socio-economic, political and demographic features of the states. For example, Scarcioffolo et al. (2018) examine socio-economic, political

[1] Identification represents the possibility of recovering a causal effect from ideal data produced by the assumed data generating process (Elwert & Winship, 2014).

and demographic determinants of the Vermont state anti-fracking bill. Meanwhile, these socio-economic, political and demographic features possibly relate to criminal activities. Assume that X is a set of observable socio-economic, political and demographic potential confounders for fracking boom D and criminal activities Y such that $D \leftarrow X \rightarrow Y$. Controlling these confounders is essential to guard against identifying spurious[2] relationships. Failure to control common cause confounders can lead to *endogeneity due to omitted variable* bias. However, over-controlling independent variables lead to loss of estimates efficiency.

Researchers do not observe the actual data generating process (DGP) and must rely on literature review and economic intuition for variables selection. However, the true DGP might involve various transformations of X, for example, lags, higher order polynomials, and interactions. Moreover, including and controlling for all these transformations may not be feasible because the covariates space grows high dimensional and regression is completely infeasible when the numbers of covariates exceed the number of observations in data. Under the assumption of sparsity (only few among these high dimensional confounders are relevant), Belloni et al. (2013) propose double-selection post-Lasso method for confounder selection (see method section).

On the other hand, confounders can be unobservable, say U. The popular causal identification strategy in this setting is the difference-in-difference (DID) model. The DID aim to attain identification by restricting the way in which unobserved confounders affect the outcome of interest over time (Abadie & Cattaneo, 2018). However, the DID identification requires an additional assumption that the unobserved confounders affect the outcome variable additively and do not change over time, which implies the "parallel trend". However, the presence of unobserved

[2] For example, even if, D does not cause Y then these two variables may be correlated because of confounding effect of X. Alternatively, if D causes X, then confounding effect of X will biased the estimates. This is known as the backdoor path criterion (Abadie & Cattaneo, 2018; Pearl, 2011).

time-varying confounders is likely to violate the "parallel trend" assumption (Xu, 2017), which leads to spurious relationships (*endogeneity via unobservable*) between treatment and outcome variables. I implement Xu (2017) Generalize Synthetic Control Method (GSC). GSC reconciles interactive fixed-effect model to absorb the unobserved time-varying confounding effect and Synthetic Control Method (SCM) as well as matrix completion method to estimate counterfactual for plausible causal interpretation.

3. Data and Empirical Strategy

3.1 Data

This study merges several panel data from various sources. The data comprises all U.S. states from 2000 to 2015. The dependent variable is the logarithmic transformation of various criminal activities per 100,000 populations and retrieved from the Federal Bureau of Investigation's (FBI) website. I merge several selected variables[3] from the University of Kentucky Center for Poverty Research (UKCPR) data; Annual State-Level Measures of Human Capital Attainment database of Frank (2009); Measures of Income Inequality database of (M. W. Frank, 2014); Top Income Shares by the State of Frank, State level employment database constructed by Barry and David was created in 2002 and is updated annually.

3.2 Double Post Lasso Selection

Consider a basic model set up:

$$y_{it} = \varphi D_{it} + \Theta'_{it}\omega + v_{it}$$

where i indexes states, t indexes times, Θ_{it} are a set of control variables to control for time-varying confounding state-level socio-economic, political and demographic features, D_{it} is a

[3] Unemployment rate, poverty rate, the fraction of state house that is the democrat, the fraction of state senate that is democrat, 2014 PCI state minimum wage (dollars), employment to population (percentage).

binary indicator if the state has fracking boom or not and y_{it} is criminal activities in each state in each time. This paper depart from the standard literature by allowing a much richer set of control variables Θ_{it} to select few proper observable confounder x_{it}. Such confounders allow controlling the possibility that some feature of a state that effects with both fracking boom and criminal activities. Therefore, after conditioning on these observable controls, exogeneity of fracking boom holds.

Causal interpretation relies on the belief that there are no higher-order terms of the control variables, no interaction terms, and no additional excluded variables that associate with the fracking boom and criminal activities. Thus, controlling a large set of variables seems desirable to make this assumption plausible. However, naively controlling redundant variables reduces the ability to distinguish the impact of interest variables and, consequently, produces less precise estimates. Belloni et al. (2013) propose the double-selection post-Lasso procedure to juggle the tradeoff between controlling for very few variables and controlling for many variables. The double-selection post-Lasso procedure is an efficient, data-driven way to search for a small set of essential confounders from among a sensibly chosen broad set of potential confounding variables. The double-post-Lasso procedure comprises the following steps:

1. In the first step, a set of control variables that are useful for predicting the feasibility of a resource boom are selected using the LASSO[4] procedure. This step helps to ensure

[4] The Least Absolute Shrinkage and Selection Operator (LASSO) is an appealing method to estimate the sparse parameter from a high-dimensional linear model is introduce by Frank & Friedman (1993) and Tibshirani (1996). LASSO simultaneously performs model selection and coefficient estimation by minimizing the sum of squared residuals plus a penalty term. The penalty term penalizes the size of the model through the sum of absolute values of coefficients. Consider a following linear model $\tilde{y}_i = \Theta_i \beta_1 + \varepsilon_i$, where Θ is high-dimensional covariates, the Lasso estimator is defined as the solution to $\min_{\beta_1 \in \mathbb{R}^p} E_n\left[\left(\tilde{y}_i -_i \Theta\beta_1\right)^2\right] + \frac{\lambda}{n}\|\beta_1\|_1$, the penalty level λ is a tuning parameter to regularize/controls the degree of penalization and to guard against overfitting. The cross-validation technique

robustness by finding control variables that are strongly related to the treatment and thus potentially important confounding factors.

2. In the second step, additional variables that predict criminal activities are selected from the control variables. This procedure ensures that important variables remain in the equation of interest, ideally helping to keep the residual variance small as well as intuitively providing an additional chance to find essential confounds.
3. In the final step, we estimate the treatment effect δ by the linear regression of y_{it} on the treatment D_{it} and the union of the set of variables selected in the two-variable selection steps.

In short, first, run Lasso of y on Θ to select a set of predictors for y. Second, run Lasso of D on Θ to select a set of predictors for D. Third, run OLS regression of y on D, and the union of the sets of regressors chosen in the two Lasso runs to estimate φ then correct the inference with the usual heteroscedasticity robust OLS standard error. For the theoretical arguments, see (Belloni et al., 2013, 2014; Nowak & Smith, 2017).

3.3 Generalized synthetic control

In this study, the outcome of interest Y_{it} is the logarithmic transformation of crime rates per 100,000 population, in the state i overtime period 1999 to 2015 indexed with a subscript t. Let τ denote the states of Arkansas, North Dakota, and West Virginia which are economically-small, relatively more rural, and have high levels of shale oil and gas reserves and c denote the state of Arizona, Georgia, Idaho, Iowa, Minnesota, Missouri, Oregon, South Carolina, South

---

chooses the best λ in prediction models. The $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$. The kinked nature of penalty function induces $\hat{\beta}$ to have many zeros, thus Lasso solution feasible model selection method.

Dakota, Washington and Wisconsin, and these states are infeasible to produce any shale gas or tight oil (Hirji & Song, 2016). The total number of states is $N = N_{tr} + N_{co}$, where N_{tr} and N_{co} are the numbers of treated and control states. Let $T_{0,i}$ be the number of pretreatment period for state i and state is first exposed to the treatment (shale boom in the year 2007) at the time $(T_{0,i} + 1)$ and observed for $q_i = T - T_{0,i}$ periods. States in the control group are never exposed to the treatment in the observed period. A linear factor model can approximate the Y_{it} and express as following form:

$$Y_{it} = \delta_{it} D_{it} + x_{it}' \beta + \lambda_i' f_t + \varepsilon_{it}$$

where, the treatment indicator is defined as D_{it} equals 1 after state i has been exposed to the treatment and equals 0 otherwise. The δ_{it} is the heterogeneous treatment effect on state i at time t; x_{it} is $k \times 1$ vector of observed confounding variables selected via double post lasso selection method; $\beta = [\beta_1, ..., \beta_k]'$ is $k \times 1$ vector of unknown parameter, $f_t = [f_{1t}, ... f_{rt}]'$ is $r \times 1$ vector of unobserved common factors that is fixed during the observed period and the treatment and control group both are affected by these fixed factors, $\lambda_i = [\lambda_{i1}, ..., \lambda_{ir}]'$ is $r \times 1$ vector of unknown factor loadings and ε_{it} represents unobserved idiosyncratic shock for state i at time t which has zero mean.

The above equation can be formalized using the potential outcome framework or Rubin's Causal Model. Let $Y_{it}(1) = x_{it}' \beta + \lambda_i' f_t + \varepsilon_{it}$ and $Y_{it}(0) = x_{it}' \beta + \lambda_i' f_t + \varepsilon_{it}$ be the potential outcome for the state i at a time t when $D_{it} = 1$ or $D_{it} = 0$ respectively. Then, the individual treatment effect on states with shale boom is as $\delta_{it} = Y_{it}(1) - Y_{it}(0)$ for any state $i \in \tau$, $t > T_0$. Then we can re-express:

$$Y_i = D_i \circ \delta_i + X_i \beta + F \lambda_i + \varepsilon_i$$

where $i \in \{1, 2, ..., N_{co}, N_{co+1}, ..., N\}$; $Y_i = [Y_{i1}, ..., Y_{iT}]'$; $D_i = [D_{i1}, ..., D_{iT}]'$; symbol $\circ$ stands for the point-wise product; $\varepsilon_i = [\varepsilon_{i1}, ..., \varepsilon_{iT}]'$; $X_i = [x_{i1}, ..., x_{iT}]'$ and $F = [f_1, ..., f_T]'$. Stacking all the control units together we can express the equation as:

$$Y_{co} = X_{co}\beta + F\lambda'_{co} + \varepsilon_{co}$$

Now to identify β, F and λ_{co}, the factors are normalized, i.e., $F'F = I_r$ and are orthogonal to each other, i.e. $\lambda'_{co}\lambda_{co} = diagonal$. These constraints are based on (Bai, 2009). Xu (2017) proposes a leave-one-out-cross-validation procedure for the choice of r or a number of factors. Then, the main quantity of interest is the average treatment effect on the treated (ATT) at the time t when $t > T_0$ and given as:

$$ATT_{t,t>T_0} = N_{tr}^{-1} \sum_{i \in \tau} \left[Y_{it}(1) - Y_{it}(0) \right] = N_{tr}^{-1} \sum_{i \in \tau} \delta_{it}$$

where, $Y_{it}(1)$ is the observed for treated units in the posttreatment period, and $Y_{it}(0)$ is the counterfactual for the treated unit in the posttreatment period. Under several assumptions[5], Xu (2017) provides the GSC estimator. In simplest, Xu (2017) the GSC estimator is a three-step process. First, the GSC estimates the interactive fixed effect model using only the control group. Second, the GSC estimates factor loadings for each treated unit by minimizing the mean squared error of the predicted treated outcome in pretreatment periods. Third, the GSC estimates counterfactuals. In practice, researchers may have limited knowledge of the exact number of factors to be included in the model. Therefore, Xu (2017) developed a cross-validation procedure to select models before estimating the causal effect. It relies on the control group information as

[5] Under the assumption of strict exogeneity (unconfoundedness), decomposable time-varying confounders, weak serial dependence of the error term, some regularity conditions and cross-sectionally independent and homoscedastic error terms.

well as information from the treatment group in pretreatment periods. Xu (2017) also incorporates the matrix completion method for causal panel data to estimate the counterfactual as shown in Athey et al. (2017). Since the counterfactual is, in a way, a prediction problem and the matrix completion from the machine learning literature imputes the missing element in a matrix, assuming the complete matrix is the sum of a low-rank matrix plus noise and the missingness is entirely at random.

The results section follows several variants of following a data generating process. This framework flexibly incorporates the additive fixed effects, known time trends, and exogenous time-invariant covariates:

$$Y_{it} = \delta_{it} D_{it} + x_{it}'\beta + \gamma_i' I_t + z_i \theta_t + \lambda_i' f_t + \alpha_i + \zeta_t + \varepsilon_{it}$$

where, Y_{it}, D_{it}, x_{it}, β, λ_i, f_t and ε_{it} holds the same interpretation as explained earlier. The I_t is a $(q \times 1)$ vector of known time trends that may affect each unit differently; γ_i is a $(q \times 1)$ unit-specific unknown parameters; z_i is a $(m \times 1)$ vector of observed time-invariant covariates; θ_t is a $(m \times 1)$ of unknown parameters, and ζ_t are additive individual and time fixed effects respectively.

4. Results and Discussion

This section highlights the main findings while supplementary results are presented in the appendix. In this study, the treatment states are West Virginia, North Dakota and Arkansas. The comparison states are shale production infeasible states, namely Arizona, Georgia, Idaho, Iowa, Minnesota, Missouri, Oregon, South Carolina, South Dakota, Washington, and Wisconsin (Hirji & Song, 2016). The logarithmic transformation of crimes per 100,000 population is the dependent variable and post 2007 is the shale boom era.

The FBI reports criminal offenses into two major categories: Part-I offenses and Part-II offenses. Part-I crimes are collectively known as Index crimes and are more severe than Part-II. The Part-I crimes include violent crimes and property crime. The violent crimes include aggravated assault, forcible rape, murder, and robbery while property crimes include burglary, larceny, motor vehicle theft, and arson.

Appendix A1 displays the list of variables, their transformation, units, and data sources. Appendix A2 shows the mean differences of various crimes. These mean differences are statistically significant. Appendix A3 shows the treatment states on average have a lower unemployment rate but are poorer, less educated with less share of the private sector but a high share of the government sector.

Table 1: Impact of shale boom on crime growth rate

Variable	DID	GSC	MCGSC
Part I Offenses			
Violent crime	0.31* (0.19)	0.46***(0.16) [0.07]	0.39** (0.18) [0.05]
Murder and non-negligent manslaughter	0.35***(0.14)	0.30 (0.27) [0.57]	0.36***(0.14) [0.20]
Forcible rape	0.53***(0.15)	1.04 (0.48) [0.13]	0.64***(0.15) [0.11]
Robbery	0.36 (0.19)	0.39** (0.21) [0.11]	0.41* (0.19) [0.10]
Aggravated assault	0.34* (0.22)	0.47***(0.17) [0.09]	0.43** (0.21) [0.05]
Property crime	0.17 (0.14)	0.12 (0.18) [0.02]	0.16 (0.11) [0.04]
Burglary	0.17** (0.08)	0.14***(0.07) [0.07]	0.15** (0.07) [0.04]
Part II Offenses			
Drug abuse violations	0.22 (0.18)	0.54** (0.18) [0.04]	0.25 (0.16) [0.05]
Embezzlement	0.70* (0.42)	0.50 (0.55) [1.45]	0.84** (0.39) [0.60]
Offenses against the family and children	-0.74***(0.33)	-0.35 (0.32) [0.09]	-0.73***(0.23) [0.09]
Prostitution and commercialize device	0.78 (0.51)	1.92 (1.07) [1.74]	0.91* (0.48) [0.55]
Vagrancy	-0.80 (0.60)	-2.86 (1.33) [0.87]	-1.23* (0.69) [0.75]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls. Appendix A4 displays full results.*

Table 1 exhibits the main findings with three different types of estimates: difference-in-difference (DID), Generalized Synthetic Control (GSC) and Matrix Completion Generalize

Synthetic Control (MCGSC). Each estimate includes the year and states the fixed effects. The double-selection post-Lasso selects the confounders. The non-parametric bootstraps (blocked at the state level) for 2,000 times provides standard errors for the DID and the MCGSC estimates. The parametric bootstraps (blocked at the state level) of 2,000 times provides standard errors for the GSC estimates.

Table 1, column (1) is a standard two-way fixed effects model often known as DID estimates in the literature. Table 1, column (2) are GSC estimates. Unlike, DID, GSC estimates impose additive state and year fixed effect and cross-validation scheme captures the unobserved factor. In GSC, the observed outcome is the sum of a linear function of covariates, and an unobserved component is a low-rank matrix plus noise. Minimizing the sum of squared errors given the rank of the matrix of unobserved components provides the estimates. Table 1, column (3) are MCGSC estimates. MCGSC is similar to GSC except it implements the matrix completion method for causal panel data to develop counterfactual shown in Athey et al. (2017).

4.1 Violent Crime

The DID estimates that violent crimes among treated groups rise at the rate of an additional[6] 0.31 compared to the comparison group. Alternatively, the DID estimates that violent crimes among treated groups is about $e^{0.31} - 1 \approx 0.36$, or 36% higher, compared to the geometric mean of violent crime among the comparison group. After conditioning on the factors, additive fixed effect,

[6] The $\ln(y) = \alpha + \delta D + \varepsilon$ shows the generic regression equation. For comparison group, the crime rises at the rate of $\alpha + \varepsilon$ i.e., $y = e^{\alpha+\varepsilon}$ if $D = 0$. For the treatment group, the crime rises at the rate of $\alpha + \delta + \varepsilon$ i.e., $y = e^{\alpha+\delta+\varepsilon}$ if $D = 1$. Alternatively, the crime in the treatment group rise by additional δ rate, compared to comparison group. The expression $\frac{e^{\alpha+\delta+\varepsilon} - e^{\alpha+\varepsilon}}{e^{\alpha+\varepsilon}} = e^{\delta+\varepsilon} - 1$ provides a percentage change analysis. Alternatively, the DID estimates that violent crimes among treated group is about $e^{0.31} - 1 \approx 0.36$, or 36% higher, compared to the geometric mean of violent crime among comparison group.

and controls, the GSC estimates that violent crime rises at an additional 0.47 rate in shale boom states compared to comparison states. The MCGSC estimates that violent crime in shale boom states rise by an additional 0.39 rate compared to the comparison states. After accounting for the unobservable, the result shows a higher impact of shale boom on violent crime compared to the DID estimate.

Figure 1: The effect of the shale boom on violent crime (1999–2015)

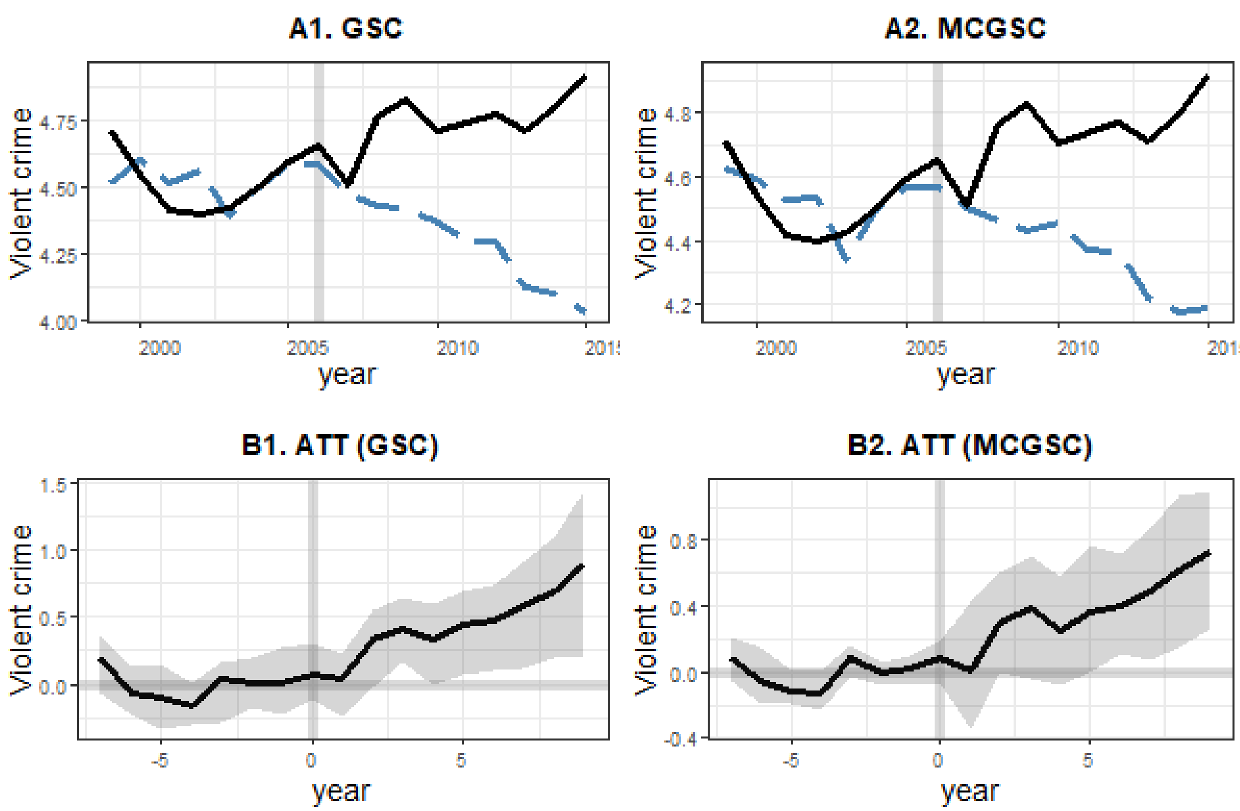


Figure 1 (panel A) shows the average actual violent crime (solid line) and average predicted violent in the absence of shale boom (dashed line); both averages are based on the number of years since (or before) shale boom or 2007. Figure 1 (panel B) shows the gap between the two lines or the estimated ATT. The confidence intervals for GSC estimates are produced by parametric block bootstraps of 2,000 times and standard error for MCGSC estimates are generated by nonparametric block bootstraps of 2000 times.

The DID estimates in Table 1, column (1) assume a constant treatment effect both across states and over time. The underlying assumption of the DID estimate is "parallel trends," which is not directly testable and less likely to hold due to unobserved heterogeneities. Figure 1 exhibits a graphical depiction that a parallel trend is plausible with GSC and MCGSC estimates for violent crimes. The Mean Squared Percentage Error (MSPE) of GSC and MCGSC are 7% and 5% respectively. Compared to GSC, MCGSC fits counterfactual better for a pretreatment outcome variable or violent crime.

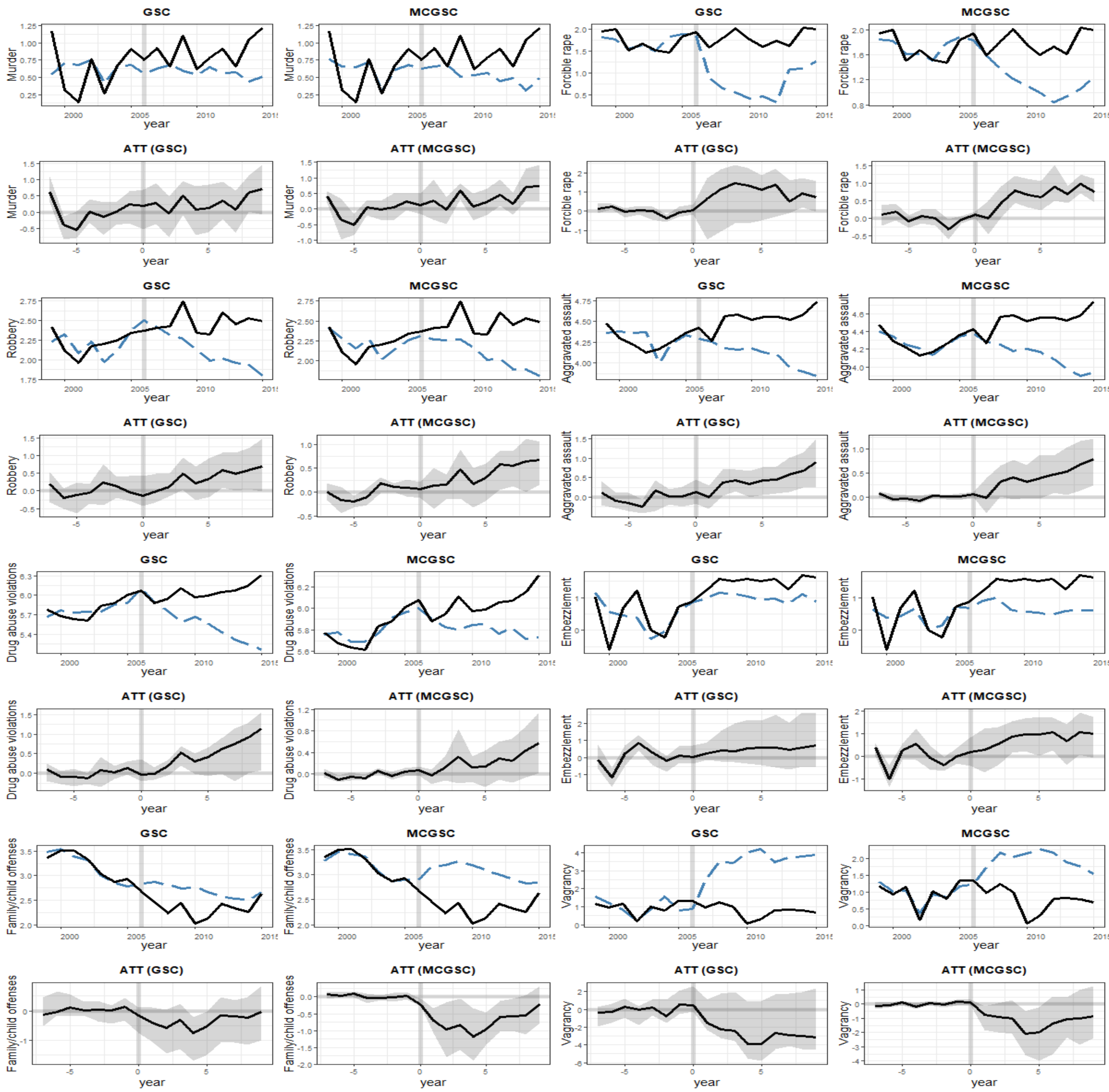
DID, GSC and MCGSC estimates show that in the shale boom states compared to the comparison states, murder and non-negligent manslaughter rose by an additional 0.35 rate; forcible rape rose by an additional 0.53 to 0.63 rate and robbery rose by a 0.40 rate and aggravated assaults by a 0.34 to 0.47 rate. Figure 2 provides a graphical representation.

4.2 Some plausible explanation of the rise in violent crime among treated states

There are possibilities of several channels to explain the increase in violent crimes during the shale boom. The fracking jobs are low-skilled and temporary jobs, mostly taken by the non-resident males who are drive-out or fly-in and fly-out employees (Ruddell et al., 2014); therefore, people with criminal records are likely to disproportionately move to areas experiencing a shale boom to find employment (James & Smith, 2017). The fracking boom is associated with an increased income inequality as the local royalty income are concentrated among a small segment of the local population; other residents do not benefit or are economically worse off (Komarek, 2018). Inequality provides a rational incentive to commit a crime (Deller & Deller, 2010). Along the same line of thought, Kelly (2000) explains that based on anomie or strain theory that inequality induces an envy effect that impacts violent crimes.

Fracking jobs are low-skilled, temporary, male-oriented jobs (Ruddell et al., 2014). Thus, the shale boom can imbalance the sex ratio and provoke specific types of crimes against women (Komarek, 2018). Businesses like a bar, prostitution and drugs boom with the fracking boom and increase illicit behavior. Residents are also disproportionately affected by disamenities like pollution, noise, water quality, and heavy traffic. This could result in tension between locals and temporary workers and, as a result, cause violent crime like aggregated assaults to rise (Komarek, 2018).

Figure 2: The effect of the shale boom on various crime (1999–2015)



Note: The average actual crime is in solid line and average predicted crime in the absence of shale boom is in dashed line; both averages are taken based on the number of years since (or before) shale boom or 2007. The gap between the two lines is the estimated ATT. The confidence intervals for GSC estimates are produced by parametric block bootstraps of 2,000 times and standard error for MCGSC estimates are produced by nonparametric block bootstraps of 2000 times

Rapid changes in economic activities and population due to the fracking boom strain the local governments, including law enforcement which could lead to ineffective policing as local governments are unable to adapt quickly to rapidly changing dynamics. For example, Kowalski & Zajac (2012) reports that the Pennsylvania State Police incidents/calls for service steadily rose for Marcellus rural counties compare to Non-Marcellus rural counties by about a third.

4.3 Property crimes and plausible explanations

Table 1 also shows DID, GSC and MCGSC estimates for property crime and various categories of property crime. Only the estimates of burglary are significant. In shale boom states compared to the comparison states; burglary is 0.14 to 0.17 higher in growth rate. Figure 2 provides a graphical representation. Komarek (2018) also finds similar results of nonexistent property-related crime. This finding counters several predictions given for rural crime literature. For example: based on social disorganization or social cohesion theory, Deller & Deller (2010) finds social capital deteriorates with the arrival of strangers, which leads to an increase in crimes. Similarly, as explained earlier, growing income disparities provides rational incentives for criminal activities (Deller & Deller, 2010), primarily via the channel defined by Kelly (2000) that anomie or strain theory predicts that inequality induces an envy effect and can lead to violent crimes related to property. Komarek (2018) explains that for property crimes to increase, the increased payoff for the crimes would have to outweigh the opportunity cost of apprehension. However, residents do not benefit or are economically worse off; therefore, property crime may not differ between boom and non-boom areas.

4.4 Part II offenses and plausible explanations

Table 1 also exhibits DID, GSC and MCGSC estimates for various categories in Part II offenses. Compared to the comparison states, GSC estimates an additional 0.54 growth rate in drug abuse violations among treated states while DID and MCGSC estimates are statistically insignificant. DID estimates an additional 0.07 growth rate, and MCGSC estimates about a 0.08 growth rate in embezzlement compared to comparison states. Figure 2 provides a graphical representation. As noted earlier, fracking jobs are low-skilled, temporary, male-oriented jobs (Ruddell et al., 2014); and people with criminal records are likely to disproportionately move to areas experiencing a shale boom to find employment (James & Smith, 2017).

However, compared to the comparison states, in shale boom states offenses against the family and children declined at the rate of 0.74 and vagrancy declined at the rate of 1.23. This result suggests a reduction in crime and possibly relates with the labor market effect that an increase in wage and employment opportunities can reduce crime by increasing the opportunity cost of apprehension. This result is consistent with Gould et al. (2002) who find increased wages, and employment for low-skilled workers reduced crime, especially among men. Similarly, Grinols & Mustard (2006) state that the resource boom leads to renovations like more street lights, the presence of residents and such developments indirectly reduces crime rates.

4.5 Falsification test

Violation of a preexisting differential trend and the sample selection of the treatment group can contaminate the estimates in this paper. The prerequisite of the study design relies upon the existence of a pretreatment parallel trend. In another words, the crime rates direction among treated and comparison groups should be statistically similar before the shale boom. To test this, I use the sample from 1999 to 2007, then I shift the fracking boom indicator variables to 2003 (4 years ahead of the actual boom). The DID, GSC and MCGSC specifications in this modified sample are statistically insignificant, which suggests a pretreatment trend does not exist. Therefore, crime rates among treatment and comparison groups are similar in the pretreatment period (See Appendix A5).

I also test whether the ruralness of treated states matters or not. Testing this effect examines how the nature of crime changes with different treatment samples of states. I consider three different samples. First, I exclude Arkansas from the main treatment sample so that a new treatment sample becomes more rural (say, group 1). Secondly, I take only Montana and Kentucky (group 2) in which the rural population is about 44.11% and 41.62% respectively. These two states are similar in terms of rural population and the way in which the shale boom mildly affects these states. Thirdly, I take the treatment group of Oklahoma, Louisiana, Michigan and New Mexico in which the rural population is 33.76%, 26.81%, 25.43%, and 22.57% respectively (group 3). Group 3 states are less rural than the previous two treatment samples. The estimates lose the statistical significance as the ruralness declines among these three treatment group states (See Appendix A6), suggesting that the shale boom increases crime among economically-small, relatively rural American states.

4.6 Robustness checks

I present the main results and falsification tests under DID, GSC, and MCGSC specifications. I use the double-selection, post Lasso method to select the contemporaneous confounding variables or controls. The estimates in the main results and falsification tests are consistent among DID, GSC, and MCGSC specifications suggesting the plausibility of robustness. However, I also proceed with two extra robustness checks. First, I allow up to two lags of control variables and select the variable with double-selection, post Lasso then I report the DID, GSC and MCGSC estimates (Appendix A7). Second, I further allow up to and as well as first-order interaction terms variables and select the variable with double-selection, post Lasso then I report the DID, GSC and MCGSC estimates (Appendix A8). The general direction of coefficients and statistical significance are consistent with the primary results in Table 1. The consistency of estimates for this two-robustness check for control specifications along with the DID, GSC, and MCGSC model specifications suggest that the results in Table 1 are likely to hold robustness.

4.7 Estimates of victimization costs of fracking

I use McCollister et al. (2010) estimates of victimization[7] cost to quantify the burden of the Part I offense crime due to the fracking boom. Instead of a log of crime rate, I use the annual frequency of crime per 100,000 residents each year and run the DID, GSC, and MCGSC models. Table 2 exhibits the estimates that are statistically significant.

Table 2: Robustness checks with first-order interaction controls and quadratic controls

Dependent variable	DID	GSC	MCGSC
Murder and non-negligent manslaughter	1.46** (0.83)	-0.42 (2.62) [4.17]	1.13** (0.53) [7.58]
Forcible rape	3.15***(0.92)	7.77 (5.13) [6.92]	3.04***(0.89) [9.02]
Robbery	5.73* (3.03)	5.68 (5.88) [16.66]	6.03** (2.53) [73.99]
Aggravated assault	26.68* (12.92)	18.61 (24.61) [874.79]	28.38** (11.58) [518.38]
Burglary	13.53** (5.57)	10.11** (5.23) [289.3]	11.79** (4.73) [172.32]
Embezzlement	1.88 (1.61)	5.24** (1.76) [8.93]	2.57 (1.74) [4.9]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls. Appendix A9 displays full results.*

Table 3 presents the annual cost of victimization among the treated states compared to the comparison states. I averaged the significant coefficients. The cost per crime is in 2008 dollars. Compared to the comparison states; the treated states suffer an extra 15.68 million dollars per year. I find that compared to the comparison states, the treated states had 1.3 more murders per 100,000

[7] The costs for each criminal category include measures of tangible costs (e.g., medical expenses, property damage) and intangible costs (e.g., elevated fear, pain, and suffering). McCollister et al. (2010) provide a comprehensive methodology for estimating the cost of violent crimes to society in 2008 dollars.

residents, suggesting about 11.63 million dollars more cost for victimization. Forcible rape appears to increase by about 3 per 100,000 residents, and victimization cost averages around 7.45 million additional dollars. Compared to the comparison states, the treated states have 27.53 more aggravated assaults which costs about an extra 2.94 million dollars. The robbery costs an extra 0.25 million dollars, and embezzlement cost about 28 thousand 2008 dollars.

Table 3: Per annum extra cost of victimization of fracking boom among treated states

Crime types	Additional crime	Cost per crime	Total cost
Murder and non-negligent manslaughter	1.295	8982907	11632864.6
Forcible rape	3.095	240776	745201.72
Robbery	5.88	42310	248782.8
Aggravated assault	27.53	107020	2946260.6
Burglary	11.81	6462	76316.22
Embezzlement	5.24	5480	28715.2
Per annum extra cost of victimization of fracking boom among treated states			**15678141.1**

Source: McCollister et al. (2010)

5. Conclusion

This paper exploits a sudden expansion of shale boom as the natural experiment to examine crime rates (Part-I and Part-II offenses as classified by Uniform Crime Report) among the relatively rural states like West Virginia, North Dakota, and Arkansas. I find consistent and robust results to support the hypothesis that the shale boom increases crime for economically-small, relatively rural American states, especially violent crime. This paper is the first to use double-selection, post lasso to control high dimensional observable controls, generalized synthetic control and matrix completion to incorporate the unobserved factors, counterfactual estimation and qualify the parallel trend assumptions for causal identification in the context of crime. These methods are based on the machine learning approach but redesigned for causal inference within the potential outcome framework. Extensive literature studies income and wage effect of the shale boom while this paper looks at the somewhat darker side of the shale boom. I find no pretreatment trend of crime, and the estimates are robust across several estimation specifications (DID, GSC, and MCGSC) as well as a model specification (lagged model, contemporaneous model, a model with polynomial and interactions) and samples. Therefore, the results support a causal interpretation. Furthermore, using McCollister et al. (2010), I find that on average each treatment state bears 15.67 million of additional victimization cost per year.

References

Abadie, A., & Cattaneo, M. D. (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, *10*(1), 465–503. https://doi.org/10.1146/annurev-economics-080217-053402

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*. https://doi.org/10.1257/000282803321455188

Allcott, H., & Keniston, D. (2018). Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies*, *85*(2), 695–731. https://doi.org/10.1093/restud/rdx042

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2017). Matrix Completion Methods for Causal Panel Data Models, 1–49. https://doi.org/10.3386/w25132

Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica*, *77*(4), 1229–1279. https://doi.org/10.3982/ECTA6135

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, *75*(2), 139–146.

Belloni, A., Chernozhukov, V., & Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608–650. https://doi.org/10.1093/restud/rdt044

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/jep.28.2.29

Deller, S. C., & Deller, M. A. (2010). Rural crime and social capital. *Growth and Change*, *41*(2), 221–275. https://doi.org/10.1111/j.1468-2257.2010.00526.x

Dobb, E. (2013). The New Oil Landscape. Retrieved January 23, 2019, from https://www.nationalgeographic.com/magazine/2013/03/bakken-shale-oil/

Eligon, J. (2013). An Oil Town Where Men Are Many, and Women Are Hounded - The New York Times. Retrieved January 23, 2019, from https://www.nytimes.com/2013/01/16/us/16women.html?pagewanted=all&_r=0

Elwert, F., & Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. *Ssrn*, (May). https://doi.org/10.1146/annurev-soc-071913-043455

Food and Water Watch. (2013). *The Social Costs of Fracking A Pennsylvania Case Study*. California. Retrieved from http://documents.foodandwaterwatch.org/doc/Social_Costs_of_Fracking.pdf

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*. https://doi.org/10.1080/00401706.1993.10485033

Frank, M. W. (2009). Inequality and growth in the united states: Evidence from a new state-level

panel of income inequality measures. *Economic Inquiry*, *47*(1), 55–68. https://doi.org/10.1111/j.1465-7295.2008.00122.x
Frank, M. W. (2014). A New State-Level panel of Annual Inequality Measures Over the Period 1916-2005. *Journal of Business Strategies*, *31*(1), 241–263.
Gould, E. D., Weinberg, B. A., & Mustard, D. B. (2002). Crime Rates and Local Labor Market Opportunities in the United States: 1979–1997. *The Review of Economics and Statistics*, *84*(1), 45–61. https://doi.org/10.1162/003465302317331919
Grinols, E. L., & Mustard, D. B. (2006). Casinos, Crime, and Community Costs. *The Review of Economics and Statistics*, *88*(1), 28–45. https://doi.org/10.1162/rest.2006.88.1.28
Grogger, J. (1998). Market Wages and Youth Crime. *Journal of Labor Economics*, *16*(4), 756–791. Retrieved from http://www.jstor.org/stable/10.1086/209905
Hirji, Z., & Song, L. (2016). Map: The Fracking Boom, State by State | InsideClimate News. Retrieved December 24, 2018, from https://insideclimatenews.org/news/20150120/map-fracking-boom-state-state
Horwitz, S. (2014). Dark side of the boom | The Washington Post. Retrieved January 23, 2019, from https://www.washingtonpost.com/sf/national/2014/09/28/dark-side-of-the-boom/?noredirect=on&utm_term=.34a122ad1ddf
James, A., & Smith, B. (2017). There will be blood: Crime rates in shale-rich U.S. counties. *Journal of Environmental Economics and Management*, *84*, 125–152. https://doi.org/10.1016/j.jeem.2016.12.004
Kelly, M. (2000). Inequality and Crime. *The Review of Economics and Statistics*, *82*(4), 530–539. https://doi.org/10.1162/003465300559028
Komarek, T. M. (2016). Labor market dynamics and the unconventional natural gas boom: Evidence from the Marcellus region. *Resource and Energy Economics*, *45*, 1–17. https://doi.org/10.1016/j.reseneeco.2016.03.004
Komarek, T. M. (2018). Crime and natural resource booms: evidence from unconventional natural gas production. *Annals of Regional Science*, *61*(1), 113–137. https://doi.org/10.1007/s00168-018-0861-x
Kowalski, L., & Zajac, G. (2012). *A preliminary examination of marcellus shale drilling activity and crime trends in Pennsylvania*.
McCollister, K. E., Frence, M. T., & Fang, H. (2010). The Cost of Crime to Society: New Crime-Specific Estimates for Policy and Program Evaluation. *Drug Alcohol Depend*, *108*, 98–109. https://doi.org/10.1016/j.drugalcdep.2009.12.002.The
Meredith, S. (2018). IEA report: Extraordinary US shale growth could force OPEC into action. Retrieved February 20, 2018, from https://www.cnbc.com/2018/02/13/iea-report-extraordinary-us-shale-growth-could-force-opec-into-action.html
Miller, K., Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746
Nowak, A., & Smith, P. (2017). Textual Analysis in Real Estate. *Journal of Applied Econometrics*, *32*(4), 896–918. https://doi.org/10.1002/jae.2550
Pearl, J. (2011). *Causality: Models, Reasoning, and Inference, Second Edition*. *Causality: Models, Reasoning, and Inference, Second Edition*. https://doi.org/10.1017/CBO9780511803161
Ruddell, R., Jayasundara, D. S., Mayzer, R., & Heitkamp, T. (2014). Drilling Down : An Examination of the Boom-Crime Relationship in Resource-Based Boom Counties. *Western*

Criminology Review.

Scarcioffolo, A. R., Shakya, S., & Hall, J. C. (2018). *The Political Economy of Vermont' s Anti-Fracking Movement*.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B*. https://doi.org/10.2307/2346178

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, *25*(1), 57–76. https://doi.org/10.1017/pan.2016.2

A1: List of Variables, details, units and data source

Variables	Details	Unit	Source
Y	Various crime rate	per 100,000	FBI
D	Shale boom	dummy	
year	Year	year id	
state	Name of the states	state id	
urate	Rate of unemployment	rate	ukcpr
pov	Rate of poverty	rate	ukcpr
govern	Dummy variable to control in the Governor is democratic	dummy	ukcpr
house	Fraction of democrats in state house	percentage	ukcpr
senate	Fraction of democrats in senate house	percentage	ukcpr
minwage	2014 PCI adjusted state minimum wage	dollars in thousands	ukcpr
emp	Employment to population ratio	percentage	ukcpr
hs	High school graduated population	percentage	Frank (2009)
col	College graduated population	percentage	Frank (2009)
atkin	Atkinson inequality coefficient	index	Frank (2014)
gini	Gini inequality coefficient	index	Frank (2014)
rmd	Root mean deviation inequality coefficient	index	Frank (2014)
thiel	Thiel inequality coefficient	index	Frank (2014)
top01us	Fraction of top 1% income population	index	Frank (2014)
millionus	Fraction of millionaires' population	index	Frank (2014)
lpcgdp	Log of per capita Gross Domestic Product (in thousands)	2014 dollars	ukcpr
lpcpi	Log of per capita income (in thousands)	2014 dollars	ukcpr
privconst	Share of private construction industry	percentage	Frank (2002)
privmanu	Share of private manufacturing industry	percentage	Frank (2002)
public	Share of total public industry	percentage	Frank (2002)
private	Share of total private industry	percentage	Frank (2002)

A2: Mean difference of dependent variable among treated and comparison states

	Treatment group		Comparison group		Diff(p-value)
Dependent variables	Mean	SD	Mean	SD	
Violent crime	4.65	0.59	4.97	0.38	-0.323***
Property crime	6.16	0.28	6.42	0.25	-0.26***
Murder and non-negligent manslaughter	0.76	0.75	1.04	0.72	-0.287**
Forcible rape	1.76	0.45	2.03	0.53	-0.266***
Robbery	2.36	0.8	3.06	0.74	-0.697***
Aggravated assault	4.42	0.63	4.67	0.39	-0.25***
Burglary	4.15	0.43	4.40	0.30	-0.252***
Larceny theft	5.94	0.30	6.2	0.25	-0.261***
Motor vehicle theft	3.07	0.36	3.41	0.56	-0.335***
Arson	1.15	0.37	1.62	0.49	-0.467***
Other assaults	5.91	0.20	6.07	0.28	-0.163***
Forgery and counterfeiting	3.37	0.56	3.5	0.60	-0.126
Fraud	4.79	0.88	4.20	0.81	0.593***
Embezzlement	1.01	1.19	1.14	1.06	-0.123
Stolen property buying receiving possessing	3.31	0.47	3.35	0.74	-0.042
Vandalism	4.23	0.30	4.67	0.45	-0.44***
Weapons carrying possessing etc.	3.50	0.49	3.82	0.52	-0.322***
Prostitution and commercialize device	1.48	1.33	2.11	1.34	-0.634***
Sex-offenses except forcible rape and prostitution	2.46	0.40	3.25	0.66	-0.795***
Drug abuse violations	5.94	0.34	6.17	0.32	-0.232***
Gambling	-0.35	1.01	-0.27	1.54	-0.08
Offenses against the family and children	2.72	0.89	3.46	0.74	-0.741***
Driving under the influence	6.26	0.34	6.29	0.34	-0.036
Liquor laws	5.38	1.01	5.80	0.65	-0.422***
Drunkenness	5.15	0.86	1.99	2.78	3.164***
Disorderly conduct	4.90	0.74	5.58	0.65	-0.673***
Vagrancy	0.86	2.03	1.43	1.68	-0.569*
All other offenses except traffic	7.12	0.60	7.03	0.48	0.089
Suspicion	0.33	1.21	0.03	1.06	0.291
Curfew and loitering law violations	2.73	1.29	2.91	1.71	-0.186
Runaways	3.02	1.44	3.86	0.80	-0.841***

A3: Mean difference of control variables among treated and comparison states

Control variables	Treatment group		Comparison group		Diff(p-value)
	Mean	SD	Mean	SD	
Unemployment rate	5.17	1.71	5.78	2.04	-0.612**
Poverty rate	14.67	3.1	12.21	2.93	2.463***
Democrat Governor	0.47	0.5	0.47	0.5	0.005
The fraction of State House that is Democrat	54.96	20.19	42.62	11.73	12.346***
Fraction of State Senate that is Democrat	59.18	20.06	43.93	13.68	15.258***
State Minimum Wage (2012 dollars)	2.94	0.23	3.09	0.44	-0.151***
Employment to Population	46.12	5.48	48.88	3.45	-2.753***
HighSchool ratio	63.44	3.44	64.18	3.49	-0.742
College ratio	15.04	3.43	18.42	2.73	-3.382***
Atkin05 index	0.25	0.02	0.27	0.02	-0.019***
Gini index	0.59	0.04	0.59	0.03	-0.003
Theil index	0.67	0.14	0.75	0.1	-0.088***
Top01us index	0.24	0.14	1.1	0.7	-0.86***
MillionUS index	0.25	0.13	1.15	0.71	-0.899***
log Per capita GDP (2012 dollars)	10.6	0.23	10.72	0.13	-0.123***
log Per capita PI (2012 dollars)	9.79	0.22	9.91	0.13	-0.123***
Share Priv Construction	5.31	0.6	5.53	1.17	-0.212*
Share Priv Manufacturing	11.17	4.29	14.13	3.47	-2.958***
share Public	19.47	2.31	16.23	1.7	3.236***

A4: Impact of shale boom on crime growth

Variable	DID	GSC	MCGSC
Part I offenses			
Violent crime	0.31* (0.19)	0.46***(0.16) [0.07]	0.39** (0.18) [0.05]
Murder and non-negligent manslaughter	0.35***(0.14)	0.30 (0.27) [0.57]	0.36***(0.14) [0.20]
Forcible rape	0.53***(0.15)	1.04 (0.48) [0.13]	0.64***(0.15) [0.11]
Robbery	0.36 (0.19)	0.39** (0.21) [0.11]	0.41* (0.19) [0.10]
Aggravated assault	0.34* (0.22)	0.47***(0.17) [0.09]	0.43** (0.21) [0.05]
Property crime	0.17 (0.14)	0.12 (0.18) [0.02]	0.16 (0.11) [0.04]
Burglary	0.17** (0.08)	0.14***(0.07) [0.07]	0.15** (0.07) [0.04]
Larceny theft	0.18 (0.17)	0.06 (0.21) [0.05]	0.17 (0.14) [0.05]
Motor vehicle theft	0.02 (0.09)	0.21 (0.32) [0.05]	0.03 (0.10) [0.05]
Arson	0.01 (0.13)	-0.11 (0.23) [0.17]	-0.01 (0.12) [0.14]
Part II offenses			
All other offenses except traffic	0.24 (0.19)	0.19 (0.27) [0.04]	0.24 (0.15) [0.07]
Curfew and loitering law violations	-0.23 (0.25)	-0.19 (0.24) [0.23]	-0.13 (0.23) [0.25]
Disorderly conduct	0.26 (0.24)	-0.05 (0.36) [0.09]	0.25 (0.18) [0.03]
Driving under the influence	-0.09 (0.13)	-0.03 (0.2) [0.04]	-0.03 (0.13) [0.06]
Drug abuse violations	0.22 (0.18)	0.54** (0.18) [0.04]	0.25 (0.16) [0.05]
Drunkenness	0.44 (0.32)	-0.01 (0.86) [0.40]	0.31 (0.34) [1.20]
Embezzlement	0.70* (0.42)	0.50 (0.55) [1.45]	0.84** (0.39) [0.60]
Forgery and counterfeiting	0.02 (0.27)	0.19 (0.16) [0.04]	0.01 (0.24) [0.06]
Fraud	-0.55 (0.33)	-0.14 (0.34) [0.16]	-0.49 (0.32) [0.08]
Gambling	-0.14 (0.20)	-0.50 (0.38) [1.05]	-0.42 (0.28) [0.67]
Liquor laws	0.07 (0.15)	0.13 (0.30) [0.10]	0.10 (0.15) [0.05]
Offenses against the family and children	-0.74***(0.33)	-0.35 (0.32) [0.09]	-0.73***(0.23) [0.09]
Prostitution and commercialize device	0.78 (0.51)	1.92 (1.07) [1.74]	0.91* (0.48) [0.55]
Runaways	-0.34 (0.39)	-0.01 (0.37) [0.28]	-0.41 (0.42) [0.17]
Sex-offenses	0.04 (0.35)	-0.41 (0.27) [0.09]	0.03 (0.3) [0.05]
Stolen property buying receiving possessing	0.26 (0.18)	0.22 (0.21) [0.15]	0.25 (0.15) [0.10]
Vagrancy	-0.80 (0.60)	-2.86 (1.33) [0.87]	-1.23* (0.69) [0.75]
Vandalism	0.16 (0.12)	0.04 (0.27) [0.02]	0.11 (0.13) [0.03]
Weapons carrying possessing etc.	0.06 (0.23)	0.20 (0.18) [0.04]	0.08 (0.21) [0.05]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls*

A5: Falsification test for 2004 pretreatment trend among treated states (1999-2007)

Dependent variable	DID	GSC	MCGSC
Violent crime	0.16 (0.25)	-0.15 (0.41) [0.17]	-0.05 (0.24) [0.06]
Property crime	0.04 (0.14)	-0.02 (0.3) [0.3]	0.00 (0.17) [0.03]
Murder and non-negligent manslaughter	0.38 (0.33)	0.07 (0.43) [0.73]	0.05 (0.32) [0.37]
Forcible rape	-0.4** (0.27)	-0.35 (0.31) [0.22]	-0.5***(0.22) [0.09]
Robbery	0.28 (0.24)	-0.15 (0.56) [0.18]	0.11 (0.3) [0.16]
Aggravated assault	0.19 (0.27)	-0.02 (0.54) [0.72]	-0.02 (0.26) [0.05]
Burglary	-0.01 (0.19)	0.45 (0.32) [0.2]	-0.04 (0.24) [0.05]
Larceny theft	0.07 (0.13)	0.01 (0.26) [0.37]	0.02 (0.16) [0.03]
Motor vehicle theft	-0.05 (0.17)	0.01 (0.36) [0.14]	-0.08 (0.2) [0.06]
Arson	-0.24 (0.25)	-0.64* (0.31) [0.38]	-0.4 (0.29) [0.15]
Other assaults	0.22 (0.12)	0.15 (0.18) [0.03]	0.15 (0.12) [0.04]
Forgery and counterfeiting	0.04 (0.18)	0.12 (0.43) [0.08]	0.00 (0.2) [0.05]
Fraud	-0.54** (0.34)	-0.43 (0.45) [0.14]	-0.75 (0.52) [0.15]
Embezzlement	0.31 (0.68)	-0.02 (0.69) [2.7]	-0.35 (0.59) [1.16]
Stolen property buying receiving possessing	0.12 (0.21)	-0.25 (0.4) [0.11]	0.07 (0.26) [0.12]
Vandalism	0.02 (0.16)	-0.05 (0.8) [0.01]	0.01 (0.18) [0.04]
Weapons carrying possessing etc.	0.18 (0.19)	0.24 (0.6) [0.01]	0.12 (0.2) [0.07]
Prostitution and commercialize device	0.15 (0.42)	0.45 (0.95) [2.22]	0.16 (0.43) [0.84]
Sex-offenses except forcible rape and prostitution	-0.03 (0.24)	-0.34 (0.42) [0.02]	-0.12 (0.23) [0.05]
Drug abuse violations	0.14 (0.17)	0.49 (0.4) [0.06]	0.11 (0.2) [0.03]
Gambling	-0.28 (0.64)	0.45 (1.02) [0.84]	-0.69 (0.71) [0.88]
Offenses against the family and children	-0.32 (0.35)	-0.28 (0.37) [0.07]	-0.40 (0.37) [0.13]
Driving under the influence	-0.01 (0.19)	-0.21 (0.18) [0.07]	0.03 (0.24) [0.08]
Liquor laws	0.32 (0.22)	0.09 (0.26) [0.09]	0.30 (0.2) [0.05]
Drunkenness	0.49 (0.57)	-0.02 (2.12) [0.47]	0.42 (0.62) [0.59]
Disorderly conduct	-0.04 (0.18)	-0.29 (0.19) [0.2]	-0.09 (0.24) [0.05]
Vagrancy	0.82 (0.52)	0.72 (2.11) [2.52]	0.78 (0.67) [0.84]
All other offenses except traffic	0.03 (0.18)	0.06 (0.26) [0.23]	-0.05 (0.2) [0.07]
Suspicion	0.67 (0.77)	-0.04 (1.04) [2.11]	-0.02 (0.85) [0.71]
Curfew and loitering law violations	0.21 (0.31)	0.07 (1.02) [0.54]	0.06 (0.41) [0.28]
Runaways	0.37 (0.23)	-0.19 (0.48) [0.44]	0.37 (0.31) [0.34]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls.*

A6: Falsification tests for ruralness among treated states

Dependent variable	Treatment group								
	A			B			C		
	DID	GSC	MCGSC	DID	GSC	MCGSC	DID	GSC	MCGSC
Violent crime	0.48***(0.19)	0.56***(0.18) [0.09]	0.54***(0.19) [0.03]	-0.36***(0.13)	-0.29* (0.16) [0.32]	-0.35***(0.14) [0.03]	0.12 (0.07)	0.17 (0.13) [0.07]	0.15 (0.07) (0.07)
Property crime	0.25** (0.15)	0.18 (0.21) [0.03]	0.25* (0.13) [0.02]	-0.1** (0.05)	-0.07 (0.14) [0.09]	-0.09** (0.05) [0.01]	0.15 (0.09)	0.19** (0.15) [0.02]	0.13* (0.08) [0.07]
Murder and non-negligent manslaughter	0.41** (0.18)	0.36 (0.33) [0.45]	0.4** (0.19) [0.15]	0.08 (0.17)	-0.04 (0.29) [0.28]	0.07 (0.17) [0.12]	-0.03 (0.13)	-0.1 (0.21) [0.38]	-0.04 (0.15) [0.17]
Forcible rape	0.53***(0.12)	1.71 (1.02) [0.15]	0.53***(0.12) [0.09]	0.08 (0.12)	0.76 (0.51) [0.09]	0.11 (0.12) [0.08]	0.06 (0.13)	0.31 (0.39) [0.04]	0.05 (0.13) [0.09]
Robbery	0.54***(0.1)	0.61** (0.25) [0.14]	0.58***(0.13) [0.06]	-0.28** (0.14)	-0.45** (0.2) [0.27]	-0.29** (0.12) [0.08]	0.04 (0.14)	0.11 (0.14) [0.11]	0.11 (0.12) [0.1]
Aggravated assault	0.44** (0.26)	0.58 (0.32) [0.08]	0.47** (0.28) [0.04]	-0.51***(0.21)	-0.43 (0.28) [0.3]	-0.48***(0.21) [0.04]	0.14 (0.09)	0.21 (0.18) [0.05]	0.17 (0.09) [0.08]
Burglary	0.21** (0.09)	0.17** (0.08) [0.04]	0.19* (0.09) [0.01]	-0.03 (0.07)	-0.06 (0.09) [0.1]	-0.02 (0.06) [0.03]	0.17** (0.06)	0.11** (0.1) [0.04]	0.15** (0.06) [0.05]
Larceny theft	0.29** (0.17)	0.13 (0.25) [0.07]	0.28* (0.16) [0.02]	-0.13** (0.05)	-0.1 (0.17) [0.13]	-0.13***(0.05) [0.02]	0.1 (0.08)	0.14* (0.1) [0.02]	0.1 (0.08) [0.07]
Motor vehicle theft	0.03 (0.1)	0.2 (0.57) [0.05]	0.01 (0.13) [0.04]	-0.23 (0.19)	-1.16 (0.67) [0.22]	-0.21 (0.2) [0.11]	0.2 (0.21)	0.25 (0.28) [0.12]	0.1 (0.22) [0.08]
Arson	0.1 (0.12)	-0.05 (0.29) [0.24]	0.08 (0.13) [0.07]	0.18 (0.12)	0.15 (0.24) [0.23]	0.18* (0.11) [0.08]	0.44** (0.3)	0.47** (0.19) [0.24]	0.44** (0.25) [0.13]
Other assaults	0.21** (0.08)	0.26 (0.2) [0.03]	0.2** (0.09) [0.01]	0.02 (0.16)	0.04 (0.16) [0.04]	0.03 (0.15) [0.02]	0.08 (0.1)	0.08 (0.12) [0.03]	0.07 (0.1) [0.05]
Forgery and counterfeiting	0.2 (0.25)	0.11 (0.23) [0.06]	0.2 (0.24) [0.04]	-0.33***(0.08)	-0.37 (0.22) [0.31]	-0.34***(0.07) [0.07]	0.02 (0.15)	-0.1 (0.15) [0.09]	-0.07 (0.15) [0.07]
Fraud	-0.28 (0.29)	0.14 (0.36) [0.09]	-0.23 (0.28) [0.07]	-0.92 (0.64)	-0.47** (0.29) [0.18]	-0.84 (0.63) [0.1]	0.25* (0.13)	0.05 (0.21) [0.03]	0.24* (0.13) [0.08]
Embezzlement	1.2***(0.57)	1.25* (0.67) [0.8]		1.15***(0.54) [0.2]			-0.05 (0.34)	-0.04 (0.56) [0.39]	0.04 (0.29) [0.23]
Stolen property buying receiving possessing	0.44***(0.13)	0.38* (0.24) [0.21]	0.43***(0.12) [0.04]	-0.18* (0.11)	-0.11 (0.53) [0.07]	-0.18* (0.11) [0.05]	-0.09 (0.15)	-0.05 (0.23) [0.11]	-0.06 (0.14) [0.12]
Vandalism	0.06 (0.12)	0.01 (0.3) [0.02]	0.03 (0.13) [0.02]	-0.25 (0.25)	-0.29 (0.25) [0.05]	-0.23 (0.25) [0.03]	0.11 (0.12)	0.03 (0.25) [0.05]	0.11 (0.12) [0.04]
Weapons carrying possessing etc.	0.26* (0.13)	0.52 (0.27) [0.04]		0.27* (0.16) [0.04]			0.14 (0.14)	0.38 (0.2) [0.04]	0.16 (0.13) [0.04]
Prostitution and commercialize device	1.08** (0.44)	1.91 (1.21) [2.26]	1.16** (0.44) [0.71]	0.1 (0.45)	0.41 (1.13) [1.85]	0.17 (0.41) [0.71]	-0.34 (0.39)	0.88 (1.12) [0.59]	-0.17 (0.43) [0.54]
Sex-offenses except forcible rape and prostitution	0.39***(0.13)	-0.13 (0.35) [0.11]	0.37***(0.15) [0.04]	-0.33 (0.39)	-0.02 (0.35) [0.46]	-0.27 (0.38) [0.07]	0.19 (0.16)	0.05 (0.22) [0.03]	0.18 (0.14) [0.06]
Drug abuse violations	0.44***(0.12)	0.82***(0.21) [0.05]	0.47***(0.11) [0.02]	0.09 (0.3)	0.59* (0.23) [0.13]	0.13 (0.29) [0.05]	0.02 (0.1)	0.2 (0.23) [0.01]	0.02 (0.09) [0.06]
Gambling	-0.21 (0.2)	-0.12 (0.42) [0.37]	-0.2 (0.2) [0.76]	-0.28 (0.96)	-0.05 (0.79) [0.67]	-0.18 (0.87) [0.94]	-0.12 (0.23)	-0.34 (0.34) [2.45]	-0.11 (0.23) [0.74]
Offenses against the family and children	-0.38** (0.2)	-0.32 (0.3) [0.08]	-0.36* (0.19) [0.11]	-0.22 (0.18)	-0.21 (0.27) [0.09]	-0.25 (0.17) [0.07]	-0.25 (0.19)	-0.04 (0.29) [0.05]	-0.27 (0.18) [0.08]
Driving under the influence	0.07 (0.1)	0.13 (0.23) [0.02]	0.09 (0.11) [0.05]	0.07 (0.12)	0.15 (0.24) [0.08]	0.09 (0.12) [0.06]	-0.16** (0.07)	-0.16 (0.17) [0.02]	-0.17** (0.08) [0.05]
Liquor laws	0.18 (0.17)	0.11 (0.37) [0.04]	0.18 (0.19) [0.03]	-0.73 (0.84)	-0.55* (0.41) [0.06]	-0.65 (0.79) [0.04]	0.06 (0.1)	0.35 (0.34) [0.17]	0.05 (0.1) [0.04]
Drunkenness	0.52 (0.34)	0.07 (0.98) [0.56]	0.45 (0.36) [0.73]	0.6* (0.33)	0.34 (0.69) [0.38]	0.55* (0.31) [0.86]	0.11 (0.35)	-0.13 (0.73) [0.29]	0.07 (0.37) [0.82]
Disorderly conduct	0.42 (0.26)	0.24 (0.44) [0.11]	0.37 (0.23) [0.03]	-0.16 (0.26)	-0.4 (0.36) [0.07]	-0.14 (0.24) [0.02]	-0.08 (0.11)	0.31 (0.36) [0.04]	-0.11 (0.12) [0.02]
Vagrancy	-0.69 (0.7)	-2.57 (1.57) [1.07]	-1.06 (0.81) [0.72]	0.73 (1.19)	-3.71 (2.15) [2.64]	0.31 (1.03) [1.13]	-0.45 (0.47)	-1.23 (0.97) [1]	-0.57 (0.47) [0.82]
All other offenses except traffic				0.2** (0.11)	0.08 (0.14) [0.21]	0.2** (0.11) [0.07]	-0.07 (0.08)	-0.09 (0.16) [0.02]	-0.07 (0.09) [0.05]
Suspicion	0.51 (0.32)	-0.17 (42.69) [0.72]	0.21 (0.24) [0.84]	0.38* (0.21)	0.11 (45.59) [0.52]	0.16 (0.15) [0.8]	0.13 (0.38)	-0.42 (13.15) [1.12]	-0.11 (0.3) [0.7]
Curfew and loitering law violations	0.15 (0.24)	-0.15 (0.35) [0.13]	0.13 (0.29) [0.43]	-1.07 (1)	-0.91***(0.21) [0.41]	-0.9 (0.98) [0.3]	-0.62** (0.28)	-0.67***(0.19) [0.44]	-0.62***(0.26) [0.2]
Runaways	0.09 (0.21)	0.3 (0.45) [0.46]	0 (0.29) [0.13]	-0.99 (0.63)	-0.9* (0.44) [0.2]	-0.93 (0.64) [0.09]	-0.44** (0.2)	-0.32 (0.32) [0.1]	-0.45** (0.19) [0.14]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls.*

A7: Robustness checks with lagged controls

Dependent variables	DID	GSC	MCGSC
Violent crime	0.52***(0.13)	0.53** (0.27) [0.05]	0.55***(0.13) [0.03]
Property crime	0.3* (0.18)	0.29***(0.14) [0.4]	0.31* (0.17) [0.01]
Murder and non-negligent manslaughter	0.37** (0.14)	0.31 (0.45) [0.26]	0.34* (0.17) [0.1]
Forcible rape	0.54** (0.25)	0.8 (0.69) [0.15]	0.54** (0.27) [0.08]
Robbery	0.59***(0.14)	0.66** (0.33) [0.08]	0.66***(0.15) [0.07]
Aggravated assault	0.56***(0.18)	0.83* (0.59) [0.1]	0.61***(0.19) [0.04]
Burglary	0.22* (0.12)	0.17** (0.09) [0.08]	0.22 (0.13) [0.01]
Larceny theft	0.34** (0.22)	0.16 (0.36) [0.5]	0.36* (0.2) [0.01]
Motor vehicle theft	0.07 (0.14)	-0.69 (0.95) [0.07]	0.04 (0.18) [0.03]
Arson	0.19 (0.22)	0.15 (0.32) [0.13]	0.17 (0.23) [0.1]
Other assaults	0.23***(0.06)	0.23 (0.22) [0.07]	0.22***(0.07) [0.02]
Forgery and counterfeiting	0.29 (0.29)	0.43* (0.23) [0.06]	0.29 (0.28) [0.02]
Fraud	-0.26 (0.33)	0.32 (0.44) [0.21]	-0.23 (0.32) [0.07]
Embezzlement	0.72** (0.4)	1.02 (0.83) [0.73]	0.81* (0.49) [0.32]
Stolen property buying receiving possessing	0.35** (0.18)	0.34 (0.45) [0.33]	0.36** (0.18) [0.05]
Vandalism	-0.05 (0.12)	-0.06 (0.22) [0.01]	-0.07 (0.15) [0.02]
Weapons carrying possessing etc.	0.31** (0.11)	0.85 (0.45) [0.04]	0.3** (0.12) [0.04]
Prostitution and commercialize device	1.11** (0.4)	5.28* (2.01) [2.65]	1.34***(0.38) [0.44]
Sex-offenses except forcible rape and prostitution	0.45***(0.11)	-0.37 (0.61) [0.04]	0.43***(0.12) [0.02]
Drug abuse violations	0.44***(0.15)	0.98** (0.32) [0.08]	0.47***(0.15) [0.04]
Gambling	-0.1 (0.38)	-0.25 (0.54) [5.71]	-0.1 (0.34) [0.91]
Offenses against the family and children	-0.26 (0.27)	0.04 (0.36) [0.05]	-0.19 (0.26) [0.08]
Driving under the influence	0.21* (0.08)	0.12 (0.36) [0.03]	0.28** (0.1) [0.03]
Liquor laws	0.31* (0.16)	0.68* (0.33) [0.11]	0.35* (0.18) [0.04]
Drunkenness	0.59 (0.42)	0.04 (1.05) [1.15]	0.46 (0.44) [1.51]
Disorderly conduct	0.47 (0.32)	0.8 (0.49) [0.15]	0.44 (0.29) [0.03]
Vagrancy	-0.54 (0.78)	-0.45 (1.88) [0.98]	-0.82 (1) [0.87]
All other offenses except traffic	0.34* (0.17)	0.43 (0.36) [0.24]	0.35* (0.16) [0.03]
Suspicion	0.45 (0.42)	-0.15 (1.08) [0.14]	0.39 (0.38) [0.7]
Curfew and loitering law violations	0.06 (0.18)	-0.02 (0.36) [0.23]	0.13 (0.21) [0.2]
Runaways	0.47 (0.36)	0.34 (0.5) [1.12]	0.42 (0.4) [0.1]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls.*

A8: Robustness checks with first-order interaction controls and quadratic controls

Dependent variables	DID	GSC	MCGSC
Violent crime	0.31* (0.19)	0.46** (0.15) [0.07]	0.39** (0.18) [0.05]
Property crime			
Murder and non-negligent manslaughter	0.35***(0.14)	0.3 (0.28) [0.57]	0.36** (0.15) [0.2]
Forcible rape	0.53***(0.15)	1.04 (0.46) [0.13]	0.64***(0.15) [0.11]
Robbery	0.36 (0.19)	0.39** (0.22) [0.11]	0.41* (0.18) [0.1]
Aggravated assault	0.34* (0.22)	0.47** (0.18) [0.09]	0.43** (0.21) [0.05]
Burglary	0.17** (0.08)	0.14***(0.07) [0.07]	0.15** (0.07) [0.04]
Larceny theft	0.18 (0.17)	0.06 (0.22) [0.05]	0.17 (0.14) [0.05]
Motor vehicle theft	0.02 (0.09)	0.21 (0.31) [0.05]	0 (0.1) [0.05]
Arson	0.01 (0.13)	-0.11 (0.24) [0.17]	-0.01 (0.12) [0.14]
Other assaults	0.29***(0.09)	0.28 (0.21) [0.04]	0.29***(0.09) [0.04]
Forgery and counterfeiting	0.02 (0.27)	0.19 (0.17) [0.04]	0.01 (0.23) [0.06]
Fraud	-0.55 (0.33)	-0.14 (0.34) [0.16]	-0.49 (0.32) [0.08]
Embezzlement	0.7* (0.42)	0.5 (0.57) [1.45]	0.84** (0.4) [0.6]
Stolen property buying receiving possessing	0.26 (0.18)	0.22 (0.21) [0.15]	0.25 (0.16) [0.1]
Vandalism	0.16 (0.12)	0.04 (0.26) [0.02]	0.11 (0.13) [0.03]
Weapons carrying possessing etc.	0.06 (0.23)	0.2 (0.19) [0.04]	0.08 (0.21) [0.05]
Prostitution and commercialize device	0.78 (0.51)	1.92 (1.04) [1.74]	0.91* (0.47) [0.55]
Sex-offenses except forcible rape and prostitution	0.04 (0.35)	-0.41 (0.28) [0.09]	0.03 (0.29) [0.05]
Drug abuse violations	0.22 (0.18)	0.54** (0.18) [0.04]	0.25 (0.16) [0.05]
Gambling	-0.48* (0.3)	-0.5 (0.38) [1.05]	-0.42 (0.28) [0.67]
Offenses against the family and children	-0.74***(0.33)	-0.35 (0.32) [0.09]	-0.73***(0.23) [0.09]
Driving under the influence	-0.09 (0.13)	-0.03 (0.18) [0.04]	-0.03 (0.13) [0.06]
Liquor laws	0.07 (0.15)	0.13 (0.31) [0.1]	0.1 (0.16) [0.05]
Drunkenness	0.44 (0.32)	-0.01 (0.81) [0.4]	0.31 (0.33) [1.2]
Disorderly conduct	0.26 (0.24)	-0.05 (0.35) [0.09]	0.25 (0.19) [0.03]
Vagrancy	-0.8 (0.6)	-2.86 (1.32) [0.87]	-1.23* (0.71) [0.75]
All other offenses except traffic			
Suspicion			
Curfew and loitering law violations	-0.23 (0.25)	-0.19 (0.24) [0.23]	-0.13 (0.24) [0.25]
Runaways	-0.34 (0.39)	-0.01 (0.4) [0.28]	-0.41 (0.42) [0.17]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls.*

A9: Impact of shale boom on crime

Dependent variable	DID	GSC	MCGSC
Violent crime	43.95** (15.72)	30.33 (34.45) [957.13]	45.4** (13.51) [1029.54]
Property crime	88.92 (59.76)	80.53* (69.68) [1578.88]	80.07* (47.39) [7818.58]
Murder and non-negligent manslaughter	1.46** (0.83)	-0.42 (2.62) [4.17]	1.13** (0.53) [7.58]
Forcible rape	3.15***(0.92)	7.77 (5.13) [6.92]	3.04***(0.89) [9.02]
Robbery	5.73* (3.03)	5.68 (5.88) [16.66]	6.03** (2.53) [73.99]
Aggravated assault	26.68* (12.92)	18.61 (24.61) [874.79]	28.38** (11.58) [518.38]
Burglary	13.53** (5.57)	10.11** (5.23) [289.3]	11.79** (4.73) [172.32]
Larceny theft	68.68 (54.42)	57.62 (58.66) [1372.65]	59.21 (42.53) [5760.54]
Motor vehicle theft	6.01 (5.34)	-3.92 (18.29) [32.59]	2.71 (5.93) [60.44]
Arson	0.85 (0.76)	0.26 (1.16) [5.83]	0.83 (0.75) [5.52]
Forgery and counterfeiting	-0.43 (13.63)	2.12 (10.44) [56.16]	-1.4 (11.8) [66.13]
Fraud	-158.05 (110.61)	-56.4 (99.44) [41721.61]	-143.4 (99.65) [7804.59]
Embezzlement	1.88 (1.61)	5.24** (1.76) [8.93]	2.57 (1.74) [4.9]
Stolen property buying receiving possessing	9.02 (6.84)	10.23 (16.45) [49.55]	6.95 (6.71) [131.54]
Vandalism	24.74 (15.64)	-10.15 (31.68) [191.03]	16.53 (16.02) [552.67]
Weapons carrying possessing etc.	4.71 (10.47)	3.37 (15.07) [50.29]	3.32 (9.36) [272.67]
Sex-offenses except forcible rape and prostitution	7.4 (5.77)	-3.27 (8.41) [8.66]	5.1 (5.29) [52.7]
Drug abuse violations	95.75 (78.2)	219.9* (94.14) [2327.23]	105.3 (83.63) [6654.03]
Gambling	0.4 (0.72)	0.31 (1.93) [1.75]	0.35 (0.73) [5.53]
Offenses against the family and children	-12.8 (9.45)	-7.6 (15.94) [192.16]	-10.84* (6.93) [142.13]
Driving under the influence	-10.31 (57.97)	36.45 (88.21) [11050.6]	20.8 (61.41) [14894.32]
Liquor laws	64.98 (86.13)	-36.09 (110.4) [6913.48]	60.33 (109.5) [12616.69]
Drunkenness	-27.37 (57.06)	-43.53 (29.89) [12664.58]	-43.68 (46.67) [1977.5]
Disorderly conduct	78.9 (87.81)	-7.39 (146.4) [1411.75]	36.93 (67.82) [20326.73]
Vagrancy	-17.58** (8.5)	-26.38 (14.18) [108.77]	-19.83** (9.26) [79.42]
All other offenses except traffic	70.81 (312.99)	87.15 (255.9) [91567.03]	88.25 (256.8) [73900.23]
Suspicion	-3.82 (6.61)	-5.07* (232.8) [382.31]	-4.46 (5.21) [109.13]
Curfew and loitering law violations	14.31 (16.56)	-9 (17.35) [187.41]	5.58 (17.57) [761.73]
Runaways	-13.34 (12.08)	-1.93 (25.2) [176.46]	-14.23 (12.91) [275.89]

*Note: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5% and 10% level of significance are given as ***, ** and * respectively. Standard errors for GSC are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The estimates in DID column is average treatment effect on treated (ATT) and the estimates in GSC and MCGSC are average of average treatment effect or AATT. All these models comprise state and year fixed effects. The controls are selected implementing double post lasso selection on the list of contemporaneous controls.*