Impact of Shale Boom on Crime among the Rural American States: Generalized Synthetic Control Approach*

Shishir Shakya†
Kazi Sohag‡

Abstract

Based on seminal and anecdotal evidence, we postulate a proposition that shale oil and gas extraction induce crime through different channels. We scrutinize the causal linkage between the fracking boom and crime rates by applying the Generalized Synthetic Control (GSC) approach in the context of Arkansas, North Dakota, and West Virginia states while considering several other states as the comparison group. We observe the prevalence of the crime rates are somewhat homogeneous before the fracking boom among treatment and comparison states or the pre-fracking boom parallel trend. And our empirical findings confirm our proposition that states with the fracking boom encountered more crimes than comparison states with an estimated 15.68 million (2008 dollar) worth of the annual victimization cost.

Keywords: Shale development; Generalize Synthetic Control; Crime; Resource curse.
JEL Classification: Q3, R1, J2

*We thanks Bob Reed, John V. Winters, Dan Rickman, Han Luyi, Alexander Lundberg, Brad Humphreys, and Joshua C. Hall

†Shippensburg University, Economics Department DHC 127, 1871 Old Main Drive, Shippensburg, PA 17257, Email: sshakya@ship.edu

‡Kazi Sohag, Graduate School of Economics Management, Ural Federal University, Russian Federation Email: ksokhag@urfu.ru

1 Introduction

There is a long debate regarding the socioeconomic cost and benefit of fracking-boom. Existing literature can be classified into two main strands stressing the causal relationship between fracking-boom and economic benefit and externalities. The first strand of literature argues that the resource boom increases wage and employment opportunities, which increases the opportunity cost of apprehension; thus, it should reduce criminal activities, (e.g., Grogger, 1998; Gould et al., 2002; Grinols and Mustard, 2006; Komarek, 2016; Allcott and Keniston, 2018; Weinstein et al., 2018; Solarin, 2020; Winters et al., 2020). In contrast, the second strand of literature focuses on the darker side of the shale boom like reductions in educational attainment (Rickman et al., 2017), adverse effects on the local area quality of life (Munasib and Rickman, 2015), sexually transmitted infections (Cunningham et al., 2020), and crime (James and Smith, 2017; Komarek, 2018; Street, 2019).

Given the existing literature on this somewhat particular question on impact of shale boom on crime, our study provides additional contributions. First, the association between the natural resource extraction boom on criminal activity may exist only among relatively rural American states. Second, we offer the victimization cost of the fracking boom. Third, we employ several novel methods with causal inference applications in the regional policy setting when treatment occurs in several states at different periods.

Our identification strategy exploits the post-2007 fracking boom period as a natural experiment. We compare various criminal activities among the shale boom states with the shale infeasible states in the pre and post-2007 periods to estimate the shale boom's impacts on various criminal activities. We consider the Part-I and Part-II offenses classified by the Uniform Crime Report as indicators of various criminal activities.

We conduct this study in relatively rural states of the USA for a few crucial reasons. The effects of the shale boom expected to be more profound at the rural state level. Besides, the influence of confounding factors tends to be less at the state level than at the county level due to heterogeneities. For example, despite all the energy development in Texas, its economy's diversity and size make it more challenging to detect energy industry impacts at the state level. Moreover, the spillover effect can make county analysis more problematic unless one accounts for the spillover in estimation. Similarly, Rickman and Wang (2020) suggest using the state-level analysis to capture spillovers, including those related to state budgets. Several research implements the state-level analysis, (e.g., Munasib and Rickman, 2015; Rickman et al., 2017; Rickman and Wang, 2020).

Following James and Smith (2017) and Komarek (2018), we consider post-2007 as the fracking boom periods. We provide a falsification test in the results section by assigning the treatment period before 2007, which concludes that consideration of post-2007 is adequate for causal analysis. We merge shale gas production data from Esri (2011), rural classification data from US Census Bureau (2012), and the fracking boom map from Hirji and Song (2016) to conclude West Virginia, North Dakota, and Arkansas as the treatment group states. These three states are more homogeneous in terms of shale production, about 50%-60% population resides in a rural area, and have experienced shale oil and gas boom only after 2007. See Table 2 for treatment states selection arguments. However, in the results section, we provide falsification tests by relaxing the treatment group selection assumptions by allowing states with fewer rural populations. The falsification test results conclude that the link between crime and shale boom is a phenomenon associated with the rural states only.

Unlike James and Smith (2017) and Komarek (2018) who perform simple difference-in-difference (DID) model with no or few covariates, we contribute by allowing high-dimensional covariates and augment DID model with several novel methods: the Belloni et al. (2014b) method of double-selection post-LASSO (DSPL) and the Xu (2017) Generalized Synthetic Control (GSC) approach. DSPL incorporates machine learning's innovation and strengths to properly select observable confounders to be partial out the causal effect of treatment effect (fracking boom) on the outcome variable (crime). Unlike Rickman and Wang (2020), Munasib and Rickman (2015), and Rickman et al. (2017)

who implement synthetic control methods that allow one treatment unit (state) at a time, GSC also allows multiple treatment units with the occurrence of treatment in a different period. Hence, this method contributes to the application side of causal inference in the regional policy setting when treatment occurs in several states at different periods.

We find a parallel trend in the pre-treatment period. The estimates are also robust across DID and GSC estimation specifications and various model specifications (lagged model, contemporaneous model, a model with polynomial, and interactions). These consistent and robust results are likely to support the hypothesis that the shale boom increases crime for relatively rural American states. Furthermore, we provide additional contribution using McCollister et al. (2010) estimates of victimization cost. On average, we find that each treatment state bears about 15.67 million additional victimization costs per year (in 2008 dollars).

Section 2 provide a brief literature review. 3 explains data, along with arguments for treatment states selection, and assumptions, strengths, and limitations of DID, DSPL, and GSC methodologies. Section 4 displays the results, discusses the mechanisms, and provides estimates of the victimization cost of the fracking boom. Section 5 concludes the study.

2 Literature Review

The seminal work of Becker (1968) on crime and punishment postulates that criminal behavior is influenced by rational incentives 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.

Extensive literature studies the income and wage effect of the shale boom. Grogger (1998) provides a theoretical prediction that the resource boom leads to an increase in wages; thus, reducing crime. Gould et al. (2002) show that the resource boom leads to increased employment and wages of low-skilled individuals; therefore, it reduces crime. Grinols and Mustard (2006) explain that the resource boom leads to renovations like more street lights and residents' presence. Thus, such developments should indirectly reduce crime rates.

A stream of studies document the positive economic impact of fracking boom (e.g., Weinstein et al., 2018; Solarin, 2020; Grinols and Mustard, 2006; Allcott and Keniston, 2018). For instance, Komarek (2016) shows evidence of growing jobs and wages in non-extraction sectors like construction, transportation, retail trade, and accommodations sector due to the fracking boom in the Marcellus region. Similarly, Allcott and Keniston (2018) find an increase in employment and wages in manufacturing. Winters et al. (2020) find positive effects of the oil and gas boom for long-term resident workers, those who were born in and reside in their home state.

In contrast, the second strand of literature focuses on the darker side of the shale boom. For example, Komarek (2018) highlights several adverse effects of the sudden shale boom in the U.S., as reported in mainstream media outlets. Rickman et al. (2017) find reductions in educational attainment in Montana, North Dakota, and West Virginia, which experienced recent significant shale-energy booms.

Munasib and Rickman (2015) find negative spillover through rising local factor and goods prices and adverse effects on the local area quality of life effect among Arkansas, North Dakota, and Pennsylvania. A recent study by Cunningham et al. (2020) find a causal relation between fracking-boom and sexually transmitted infections (STI) rate in fracking counties. The study argues that high-fracking production areas encounter a higher rate of prostitution due to male in-migration.

James and Smith (2017) quantify the cost of shale boom-related crimes for the USA in aggregate level and associate increase in the young male population among shale boom-counties to criminal activity. Similarly, Komarek (2018) studies the counties in Pennsylvania and New York, and Street (2019) study counties of North Dakota. They both find a link between fracking and crime. Dobb (2013) argue fracking-boom induced resettlement increases binge drinking, prostitution, and violence in North Dakota. Anecdotal evidence shows that due to fracking-boom, drug-related crimes surged significantly in Washington (Horwitz, 2014). Similarly, the New York Times reports increased violence against women (Eligon, 2013), fatal accidents, and sexually transmitted infections (Food and Water Watch, 2013). The ongoing debate motivates us to re-investigate the casual effect of fracking-boom on criminal activities.

3 Data and Empirical Strategy

3.1 Data

This study utilizes 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 offenses per 100,000 populations and retrieved from the Federal Bureau of Investigation's (FBI) website. We merge several selected variables from the University of Kentucky Center for Poverty Research data (UKCPR, 2018); Annual State-Level Measures of Human Capital Attainment database (Frank, 2009); Measures of Income Inequality database (Frank, 2014); State-level employment database (Hirsch and MacPherson, 2003). Table 1 provides the list of variables, their respective details, units, and data sources.

In Table 2, We merge Esri (2011) shale production data with US Census Bureau (2012) data on the percentage of the population residing in the rural area data, and include only shale producing states. We can promptly see that West Virginia, North Dakota, and Arkansas are relatively rural states and have experienced a moderate shale boom. Oklahoma and Michigan have also experienced a moderate amount of shale boom but are relatively less rural.

States not included in Table 2 are either shale infeasible states or feasible but non-shale producing states. Hirji and Song (2016) provides a list of states that are not economically feasible for shale production. Based on Hirji and Song (2016). We include the states that

Table 1: 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)

Notes: In the regression, we do not include all of these variables. We employ double-selection post-LASSO method for variable selection.

are not economically feasible for shale production as the control group. These states are Arizona, Connecticut, Delaware, Georgia, Idaho, Iowa, Maine, Minnesota, Missouri, New Hampshire, New Jersey, Oregon, Rhodes Island, South Carolina, South Dakota, Washington, and Wisconsin.

3.2 Double Selection Post-LASSO

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 $D \longrightarrow Y$. It is improbable that crime affects the fracking boom; hence the endogeneity from reverse causality $D \longleftarrow Y$ is doubtful. The availability of shale reserves is truly exogenous; however, after the fracking innovation is technologically feasible, the fracking boom might occur with political and business mo-

Table 2: Treatment states: rural and moderate shale boom states, sorted by rural population

States	Rural population (%)	Shale production (annual 2011)
West Virginia[§]	51.28	Moderate
Montana	44.11	Low
Arkansas[§]	43.84	Moderate
Kentucky	41.62	Low
North Dakota[§]	40.10	Moderate
Wyoming[*]	35.24	Moderate
Oklahoma	33.76	Moderate
Louisiana	26.81	High
Michigan	25.43	Moderate
New Mexico	22.57	Low
Pennsylvania	21.34	High
Texas	15.30	High
Colorado	13.85	Low
Puerto Rico	6.24	Moderate
California	5.05	Moderate
District of Columbia	0.00	Moderate

Notes: Marked with [§] are the treatment states. US Census Bureau (2012) provides state-level data on the percentage of the population residing in rural areas. Esri (2011) provides the list of shale gas production by state in 2011 and categorizes states producing more than 1,000 billion cubic feet (bcf) per year, in between 300-1,000 bcf, below 300 bcf as High, Moderate, and Low respectively. The states not included are shale infeasible, and non-shale producing states. Wyoming is marked with [*]. Wyoming is a major fracking state (especially in per capita terms), with more wells drilled than treatment states AR and WV (Ridlington et al., 2016), but has rural population of 35.24%. I have exclude Wyoming in the main analysis but included in Appendix A Table A4.
Source: https://storymaps.esri.com/stories/2013/ShaleGas/
Source: https://insideclimatenews.org/news/20150120/map-fracking-boom-state-state

tives, which possibly depend upon the socioeconomic, political, and demographic features of the states.[1] Assume that X is a set of observable socioeconomic, political, and demographic, potential confounders and covariates for fracking boom D and criminal activities Y such that $D \longleftarrow X \longrightarrow Y$. 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

[1]For example, Scarcioffolo et al. (2020) examine socioeconomic, political, and demographic determinants of the Vermont state anti-fracking bill. Meanwhile, the literature points out that these socioeconomic, political, and demographic features possibly relate to criminal activities.

[2]For example, even if D does not cause Y, these two variables may be correlated because of the confounding effect of X. Alternatively, if D causes Y, then the confounding impact of X will bias the estimates. This is known as the backdoor path criterion (Abadie and Cattaneo, 2018; Pearl, 2009).

variables lead to loss of estimates efficiency. However, 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. Under the assumption of sparsity (only a few among these high dimensional confounders are relevant), Belloni et al. (2014b) propose double-selection post-LASSO method for confounder selection using a basic difference-in-difference (DID) set up:

$$Y_{it} = \varphi D_{it} + \Theta'_{it}\omega + \nu_{it}$$

where i indexes state, t indexes times, Θ_{it} are a set of state-level socioeconomic, political, and demographic features, D_{it} is a binary indicator if the state has a fracking boom or not, and Y_{it} is criminal activities in each state in each time. This paper departs from the standard literature by allowing a much richer set of control variables to select a few proper observables $x_{it} \in \Theta_{it}$. We assume the unconfoundedness assumption that exogeneity of fracking boom holds after adequately controlling features of a state that affects both fracking boom and criminal activities or $Y_{it} \perp D_{it}|x_{it}$.

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 associated 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. (2014b) 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. Belloni et al. (2014a) explains the double-post-LASSO procedure as following two LASSO[3] steps:

[3]The Least Absolute Shrinkage and Selection Operator (LASSO) is an appealing method to estimate

- In the first step, a set of control variables that are useful for predicting the feasibility of a resource boom, D, are selected using the LASSO procedure. This step helps to find control variables that are strongly related to the treatment (fracking boom). If the treatment is genuinely exogenous, then this procedure should select no variables.

- In the second step, variables that predict criminal activities, Y, are selected from the control variables using the LASSO procedure. This procedure ensures that essential variables remain in the equation of interest, ideally helping to keep the residual variance small and intuitively providing an additional chance to find essential confounds.

- 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 that are selected in the previous two steps. This procedure is known as double-selection. Finally, we allow heteroscedasticity robust OLS standard error for correct inference, known as a post-LASSO estimation.

3.3 Generalized synthetic control

The DSPL allows a proper selection of observables; however, confounders can be unobservable. This setting's popular causal identification strategy is the difference-in-difference (DID) model and event study methodology. The DID and event study aim to attain identification by restricting how unobserved confounders affect the outcome of interest over time (Abadie and Cattaneo, 2018). However, the DID identification requires an additional assumption that the unobserved confounders affect the outcome variable

---

the sparse parameter from a high-dimensional linear model and is introduced 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 $\min_{\beta_1 \in \mathbb{R}^p} E_n \left[(\tilde{y}_i - \Theta_i\beta_1)^2\right] + \frac{\lambda}{n}\|\beta_1\|_1$, the penalty level λ is a tuning parameter to regularize/controls the degree of penalization and to guard against overfitting. The cross-validation technique chooses the best λ in prediction models and $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$. The penalty function's kinked nature induces $\hat{\beta}$ to have many zeros; thus, the LASSO solution is feasible for model selection.

additively and do not change over time, which implies the "parallel trend." The event study enables us to detect a parallel trend in the pre-treatment period visually. A better identification strategy against DID and event study is the Abadie et al. (2010) synthetic control (S.C.) method. The SC method can account for the effects of confounders changing over time by weighting the control group to better match the treatment group before the intervention, which enables researchers to systematically select comparison groups in such a way that the control group mirrors the treatment group as closely as possible in pre-treatment period.

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

GSC is a relatively novel method and advances the DID literature. First, DID is a particular case of Generalized Synthetic Control (GSC). Second, unlike the S.C. approach, GSC can allow multiple treatment units with treatment at a different time. Third, SC uses only positive weights on the control group's donor pool to mirror the treatment unit. However, GSC allows both positive and negative weights to reproduce treatment variables, which allows a better chance to capture the pre-treatment parallel trend if any exits. Fourth, DID, SC, and event studies usually incorporate time and unit fixed effects, generally known as two-way fixed effects. These fixed effects allow absorbing the shocks whose impact is restricted to a given year and county, respectively. The interactive fixed effect is augmented in GSC. The interactive fixed effects allow modeling time-varying unobserved heterogeneities explicitly. Usually, quadratic or cubic trend terms are augmented to standard DID to model the time-varying unobserved heterogeneities. Simultaneously, the GSC allows interactive fixed-effect models to model the non-linearities of time-varying unobserved heterogeneities explicitly. These additional properties of the GSC method make GSC more plausible than standard difference-in-differences assumptions. However, the foundation of GSC begins with standard assumptions of DID. Therefore, in my analysis, we have included DID estimates, along with GSC.

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 relatively more rural and have experienced moderate levels of shale oil and gas reserves, and c denote the state of Arizona, Connecticut, Delaware, Georgia, Idaho, Iowa, Maine, Minnesota, Missouri, New Hampshire, New Jersey, Oregon, Rhodes Island, South Carolina, South Dakota, Washington, and Wisconsin. These states are infeasible to produce any shale gas or tight oil.

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 pre-treatment 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 the 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 observables selected via DSPL 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 $Y_{it} = \delta_{it}D_{it}+x'_{it}\beta+\lambda'_{i}f_{t}+\varepsilon_{it}$ and $Y_{it} = \delta_{it}D_{it}+x'_{it}\beta+\lambda'_{i}f_{t}+\varepsilon_{it}$ 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 $\delta_{it} = Y_{it}\left(1\right) - Y_{it}\left(0\right)$ 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 the Bai (2009) model. 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:

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

where, $Y_{it}\left(1\right)$ is the observed for treated units in the post-treatment period, and $Y_{it}\left(0\right)$is the counterfactual for the tree post-treatment period. Under several assumptions[4] Xu (2017) provides the GSC estimator.

In simplest, Xu (2017) the GSC estimator is a three-step process. First, the GSC

[4]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.

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 pre-treatment periods. Third, the GSC estimates counterfactuals. In practice, researchers may have limited knowledge of the exact number of factors 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 well as information from the treatment group in pre-treatment periods.

Now, we can extend the above modeling framework to a more general difference-in-difference model that flexibly incorporates the additive fixed effects and interactive fixed-effects along with covariates as:

$$Y_{it} = \delta_{it} D_{it} + x'_{it}\beta + \lambda'_i f_t + \alpha_i + \zeta_t + \varepsilon_{it}$$

where, Y_{it}, D_{it}, x_{it}, β, λ_i, f_t and ε_{it} holds the same interpretation as explained earlier. The α_i and ζ_t are additive individual and time fixed effects respectively.

The fundamental problem of causal inference is that the treated and control group's counterfactuals are unknown to the researcher; therefore, half of the data is always missing. Athey et al. (2018) assumes that the counterfactual is, in a way, a prediction problem and uses the matrix completion from the machine learning literature imputes the missing element in a matrix. Unlike DID, GSC estimates impose additive state and year fixed effect, and the 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. Xu (2017) also incorporates the matrix completion method for causal panel data to estimate the counterfactual, as shown in Athey et al. (2018).

4 Results and Discussion

As explained earlier, my identification strategy exploits the post-2007 fracking boom period as a natural experiment. It compares various criminal activities of shale boom states with the shale infeasible states to estimate the shale boom's impacts on various criminal activities. Following James and Smith (2017) and Komarek (2018), I consider post-2007 as the fracking boom periods. Merging shale gas production data Esri (2011), rural classification data US Census Bureau (2012), and the fracking boom map Hirji and Song (2016), I conclude West Virginia, North Dakota, and Arkansas as the treatment group states. These states are relatively rural and have experienced the shale oil and gas boom only after 2007. 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. Part-I crimes include violent crimes and property crimes. Violent crimes include aggravated assault, forcible rape, murder, and robbery. In contrast, property crimes include burglary, larceny, motor vehicle theft, and arson. Table 1 displays the variables, transformation, units, and data sources.

This section highlights the main findings, while supplementary results are presented in Appendix A. Appendix A, Table A1, panel A shows the mean differences of various crimes. Panel B shows the mean differences of various control variables among treated and control groups. Most of these mean differences are statistically significant. The treatment states, on average, have a lower unemployment rate. Still, they are poorer, less educated, with less share of the private sector but a high share of the government sector.

Table 3 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 nonparametric bootstraps (blocked at the state level) for 2,000 times provide standard errors for the

Table 3: Impact of shale boom on crime growth rate

Variable	DID (1)	GSC (2)	MCGSC (3)
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]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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) 2,000 times. The estimates in the DID column is the 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 the double-selection post-LASSO on the list of contemporaneous controls. Appendix A. Table A2 displays full results.

DID and the MCGSC estimates. The parametric bootstraps (blocked at the state level) of 2,000 times provide standard errors for the GSC estimates. Table 3, column (1), is a standard two-way fixed effects model, often known as DID estimates in the literature. Table 3, column (2) are GSC estimates. Unlike DID, GSC estimates impose additive state and year fixed effect, and the 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 3, 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. (2018).

4.1 Violent crime

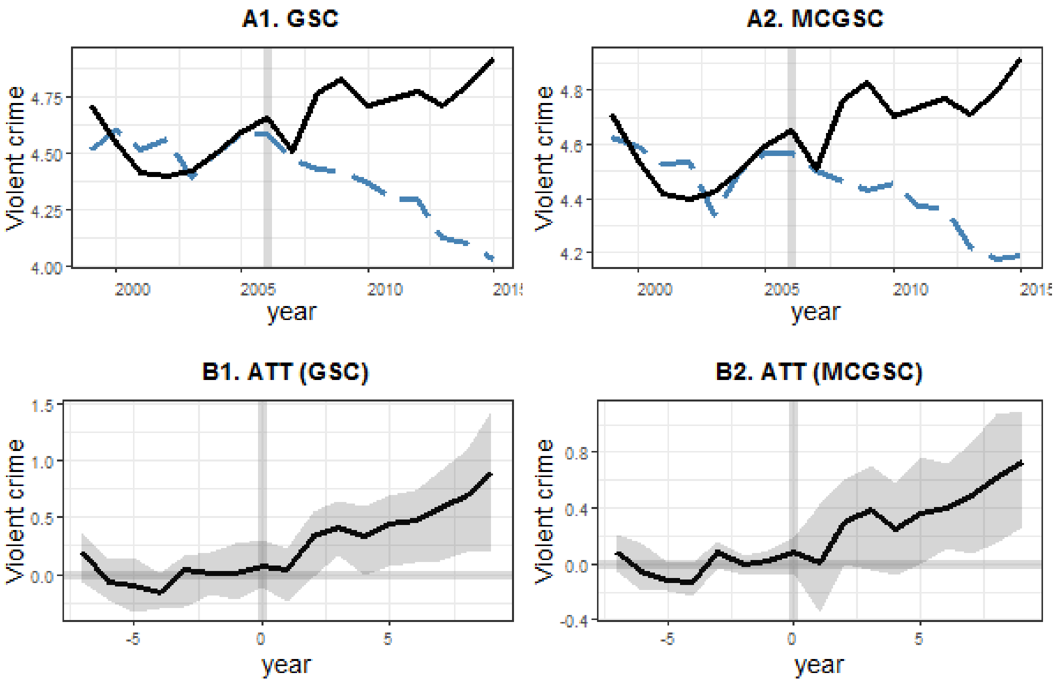
The DID estimates that violent crimes among treated groups rise at the rate of an additional[5] 0.31 compared to the comparison group. Alternatively, the DID estimates that violent crimes among treated groups are about $e^{0.31} - 1 \approx 0.36$, or 36% higher, compared to the geometric mean of violent crime among the comparison group. After conditioning the factors, additive fixed effect, and controls, the GSC estimates that violent crime rises at an additional 0.47 rate in shale boom states than in comparison states. The MCGSC estimates that violent crime in the shale boom states rises by an additional 0.39 rate than the comparison states. After accounting for the unobservable, the result shows a higher impact of the shale boom on violent crime than the DID estimate.

The DID estimates in Table 3, 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 pre-treatment 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 a further 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.

[5] $\ln(y) = \alpha + \delta D + \varepsilon$ 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 treatment group's crime rises by an additional δ rate than the 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 the treated group are about $e^{0.31} - 1 \approx 0.36$, or 36% higher, compared to the geometric mean of violent crime among the comparison group.

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



Notes: 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.

4.2 Some plausible explanation of the rise in violent crime

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, 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 and Smith, 2017). The fracking boom is associated with increased income inequality as the local royalty income is 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 and Deller, 2010). Along the same line of thought, Kelly (2000) explains that inequality induces an envy effect that impacts violent crimes based on anomie or strain theory.

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).

Rapid changes in economic activities and population due to the fracking boom strain the local governments, including law enforcement, leading to ineffective policing as local governments cannot adapt quickly to rapidly changing dynamics. For example, Kowalski and 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 3 also shows DID, GSC, and MCGSC estimates for property crime and various categories of property crime. Only the estimates of burglary are significant. Compared to the comparison states, in shale boom states, burglary is 0.14 to 0.17 higher 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 and 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 provide rational incentives for criminal activities (Deller and 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 3 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. In contrast, DID and MCGSC estimates are statistically insignificant. DID estimates an additional 0.07 growth rate, and MCGSC estimates 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 and 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 dropped at 1.23. This result suggests a crime reduction and possibly relates to the labor market effect. An increase in wages 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 decreased crime, especially among men. Similarly, Grinols and Mustard (2006) states that the resource boom leads to renovations like more street lights, residents' presence, and such developments indirectly reduce crime rates.

4.5 Falsification test

Violation of a pre-treatment 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 pre-treatment parallel trend. In other words, the crime rate direction among treated and comparison groups should be statistically similar before the shale boom. To test this, we use the sample from 1999 to 2007, and 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, suggesting a post-treatment effect on crime does not exist. And, crime rates among treatment and comparison groups are similar in the pre-treatment and post-treatment period (See Appendix A, Table A3). Therefore, the choice of 2007 as the shale boom seems appropriate.

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 primary treatment sample so that a new treatment sample becomes more rural (say, group 1). Secondly, I take only Montana, Kentucky, and Wyoming (group 2), in which the rural population is about 44.11%, 41.62%, and 35.24%, respectively. These states are similar in terms of the rural population and how they mildly affect 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.

4.6 Robustness checks

We present the main results and falsification tests under DID, GSC, and MCGSC specifications. We use the DSPL 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 robustness plausibility. However, I also proceed with two extra robustness checks. First, we allow up to two lags of control variables and select the variable with DSPL. Then I report the DID, GSC, and MCGSC estimates (Appendix A. Table A5). Second, we further allow up to and as well as first-order interaction terms variables and select the variable with DSPL. Then we report the

DID, GSC, and MCGSC estimates (Appendix A. Table A6). The general direction of coefficients and statistical significance are consistent with the initial results in Table 3. 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 3 are likely to hold robustness.

4.7 Estimates of victimization costs of fracking

I use McCollister et al. (2010) estimates of victimization[6] 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 4 exhibits that such incidence rates are statistically significant and non zero.

Table 5 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 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 additional 0.25 million dollars, and embezzlement cost about 28 thousand 2008 dollars.

[6]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) provides a comprehensive methodology for estimating the cost of violent crimes to society in 2008 dollars.

Table 4: Additional crime among treated states

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]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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) 2,000 times. The DID column estimates are 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 the double-selection post-LASSO on the list of contemporaneous controls. Appendix A, Table A7 displays full results.

Table 5: 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	11632865
Forcible rape	3.095	240776	745201.7
Robbery	5.88	42310	248782.8
Aggravated assault	27.53	107020	2946261
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**

Sources: McCollister et al. (2010)

5 Conclusion

This paper exploits a sudden expansion of the shale boom as a natural experiment to examine crime rates among the relatively rural states like West Virginia, North Dakota, and Arkansas. We find consistent and robust results to support the hypothesis that the shale boom increases crime for 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 the income and wage effect of the shale boom. At the same time, this paper looks at the somewhat darker side of the shale boom. We find a pre-treatment parallel trend of crime. The estimates are robust across several estimation specifications (DID, GSC, and MCGSC) and a model specification (lagged model, contemporaneous model, a model with polynomial and interactions) and samples. Therefore, the results are likely to support a causal interpretation.

Furthermore, using (McCollister et al., 2010), We find that, on average, each treatment state bears 15.67 million of additional victimization cost per year. Our empirical findings suggest that policymakers should consider the higher relevance of the crime rate as a potential cost to the fracking boom and strengthen the public services, including law and order. Finally, the distributional aspect should be improved to enhance the welfare of dwellers.

References

Abadie, A. and Cattaneo, M. D. (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, 10(1):465–503.

Abadie, A., Diamond, A., and 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.

Allcott, H. and 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.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2018). Matrix Completion Methods for Causal Panel Data Models. Working Paper 25132, National Bureau of Economic Research.

Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279.

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

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-dimensional methods and inference on structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014b). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650.

Cunningham, S., DeAngelo, G., and Smith, B. (2020). Fracking and Risky Sexual Activity. *Journal of Health Economics*, page 102322.

Deller, S. C. and Deller, M. A. (2010). Rural crime and social capital. *Growth and Change*, 41(2):221–275.

Dobb, E. (2013). The New Oil Landscape.

Eligon, J. (2013). An Oil Town Where Men Are Many, and Women Are Hounded - The New York Times.

Esri (2011). Mapping the Shale Gas Boom - A Story Map Presented by Esri.

Food and Water Watch (2013). The Social Costs of Fracking A Pennsylvania Case Study. Technical report, Food & Water Watch, California.

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.

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., and 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.

Grinols, E. L. and Mustard, D. B. (2006). Casinos, Crime, and Community Costs. *The Review of Economics and Statistics*, 88(1):28–45.

Grogger, J. (1998). Market Wages and Youth Crime. *Journal of Labor Economics*, 16(4):756–791.

Hirji, Z. and Song, L. (2016). Map: The Fracking Boom, State by State — Inside Climate News.

Hirsch, B. T. and MacPherson, D. A. (2003). Union Membership and Coverage Database from the Current Population Survey: Note. *Industrial and Labor Relations Review*, 56(2):349–54.

Horwitz, S. (2014). Dark Side of the Boom — The Washington Post.

James, A. and Smith, B. (2017). There will be Blood: Crime Rates in Shale-rich U.S. Counties. *Journal of Environmental Economics and Management*, 84:125–152.

Kelly, M. (2000). Inequality and Crime. *The Review of Economics and Statistics*, 82(4):530–539.

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.

Komarek, T. M. (2018). Crime and Natural Resource Booms: Evidence from Unconventional Natural Gas Production. *Annals of Regional Science*, 61(1):113–137.

Kowalski, L. and Zajac, G. (2012). A Preliminary Examination of Marcellus Shale Drilling Activity and Crime Trends in Pennsylvania. Report, Justice Center for Research.

McCollister, K. E., Frence, M. T., and Fang, H. (2010). The Cost of Crime to Society: New Crime-Specific Estimates for Policy and Program Evaluation. *Drug Alcohol Depend*, 108:98–109.

Munasib, A. and Rickman, D. S. (2015). Regional Economic Impacts of the Shale Gas and Tight Oil Boom: A Synthetic Control Analysis. *Regional Science and Urban Economics*, 50:1–17.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Rickman, D. and Wang, H. (2020). What Goes up must Come Down? The Recent Economic Cycles of the Four most Oil and Gas Dominated States in the US. *Energy Economics*, 86:104665.

Rickman, D. S., Wang, H., and Winters, J. V. (2017). Is Shale Development Drilling Holes in the Human Capital Pipeline? *Energy Economics*, 62:283 – 290.

Ridlington, E., Norman, K., and Richardson, R. (2016). Fracking by the Numbers The Damage to Our Water, Land and Climate from a Decade of Dirty Drilling.

Ruddell, R., Jayasundara, D. S., Mayzer, R., and 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., and Hall, J. C. (2020). The Political Economy of Vermont's Anti-fracking Movement. *Journal of Public Finance and Public Choice*.

Solarin, S. A. (2020). The Effects of Shale Oil Production, Capital and Labour on Economic Growth in the United States: A Maximum Likelihood Analysis of the Resource Curse Hypothesis. *Resources Policy*, 68:101799.

Street, B. (2019). The Impact of Economic Opportunity on Criminal Behavior: Evidence from the Fracking Boom. Working Paper, The Center for Growth and Opportunity at Utah State University.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso.

UKCPR (2018). National Welfare Data — University of Kentucky Center for Poverty Research.

US Census Bureau (2012). 2010 Census Urban and Rural Classification and Urban Area Criteria. *census.gov*.

Weinstein, A. L., Partridge, M. D., and Tsvetkova, A. (2018). Follow the Money: Aggregate, Sectoral and Spatial Effects of an Energy Boom on Local Earnings. *Resources Policy*, 55:196–209.

Winters, J. V., Cai, Z., Maguire, K., and Sengupta, S. (2020). Causal Effects of the Fracking Boom on Long-Term Resident Workers. *Journal of Regional Science*.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, 25(1):57–76.

A Appendix A

Table A1: Mean differences among treated and comparison states

Variables	Treatment group		Comparison group		Diff(p-value)
	Mean	SD	Mean	SD	
Panel A, Dependent variables					
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	4.4	0.3	-0.252***
Larceny theft	5.94	0.3	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	0.2	6.07	0.28	-0.163***
Forgery and counterfeiting	3.37	0.56	3.5	0.6	-0.126
Fraud	4.79	0.88	4.2	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	0.3	4.67	0.45	-0.44***
Weapons carrying possessing etc.	3.5	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	0.4	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	5.8	0.65	-0.422***
Drunkenness	5.15	0.86	1.99	2.78	3.164***
Disorderly conduct	4.9	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	0.6	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	0.8	-0.841***
Panel B, Control variables					
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***

Table A2: Impact of shale boom on crime growth

Variable	DID (1)	GSC (2)	MCGSC (3)
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]

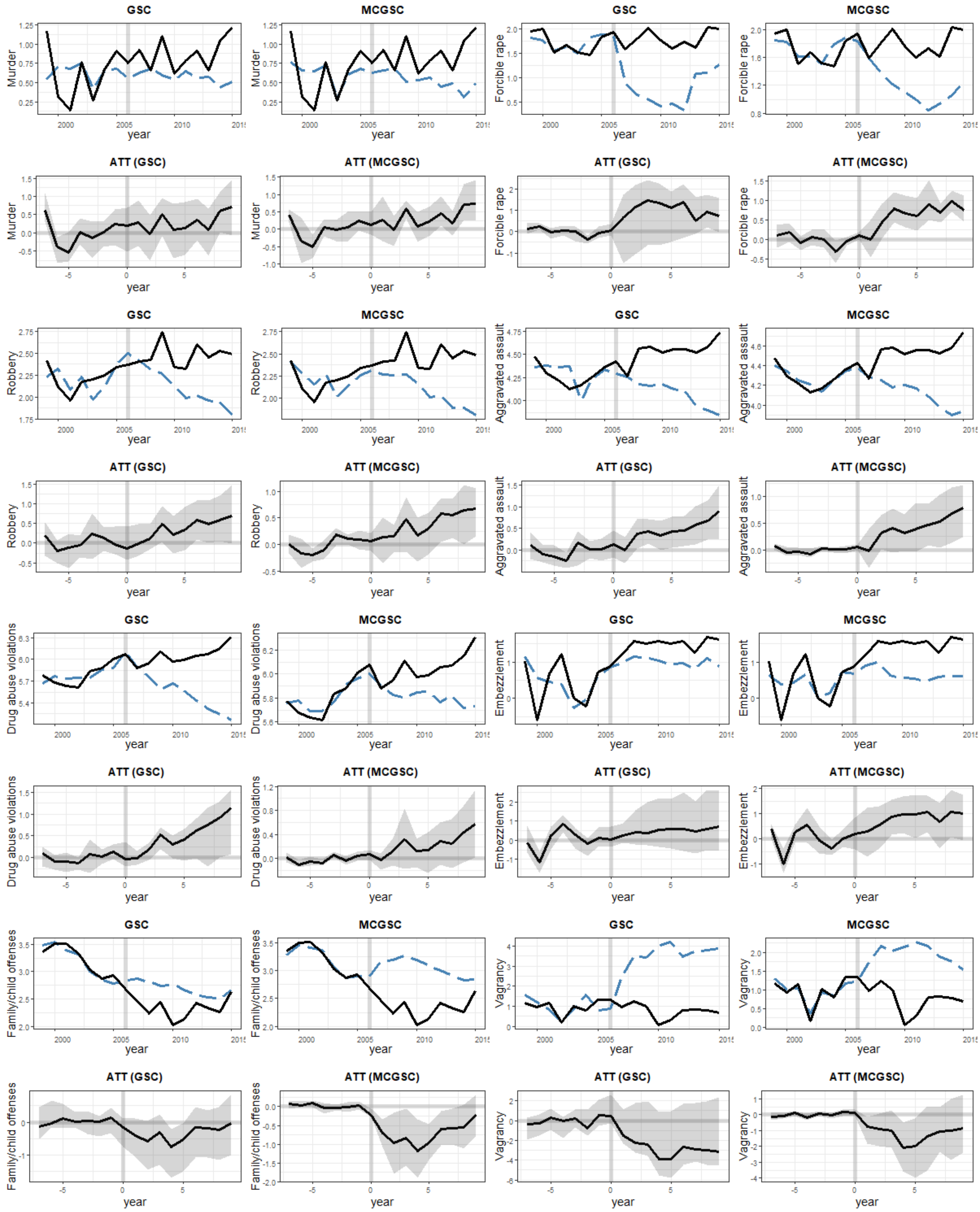
Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors are based on parametric bootstraps (blocked at the state level) of 2,000 times. Standard errors for DID and MCGSC are based on nonparametric bootstraps (blocked at the state level) of 2,000 times. The DID column estimates are 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-selection post-lasso on the list of contemporaneous controls.

Table A3: Falsification test for 2004 pre-treatment 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]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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 DID column estimates are 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-selection post-lasso on the list of contemporaneous controls.

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



Notes: 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.

Table A4: 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]	-0.0375	-0.18 (0.15) [0.16]	-0.23** (0.13) [0.06]	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]	-0.06 (0.06)	-0.03 (0.12) [0.07]	-0.06 (0.06) [0.05]	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]	0.1 (0.11)	-0.18 (0.31) [0.21]	0.08 (0.11) [0.22]	-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]	0.15 (0.12)	0.68 (0.46) [0.09]	0.18* (0.11) [0.11]	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]	-0.27***(0.1)	-0.37* (0.18) [0.21]	-0.27***(0.09) [0.11]	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]	-0.0735	-0.33 (0.18) [0.16]	-0.32* (0.19) [0.07]	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]	-0.07 (0.06)	-0.08 (0.07) [0.07]	-0.06 (0.06) [0.04]	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]	-0.07 (0.06)	-0.04 (0.15) [0.09]	-0.07 (0.06) [0.06]	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]	-0.04 (0.21)	-0.4 (0.48) [0.2]	-0.04 (0.21) [0.07]	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]	0.2* (0.11)	-0.03 (0.29) [0.33]	0.18* (0.11) [0.12]	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]	0.04 (0.12)	0.07 (0.14) [0.08]	0.05 (0.11) [0.04]	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]	-0.28***(0.1)	-0.33 (0.19) [0.12]	-0.29***(0.09) [0.08]	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]	-0.48 (0.56)	-0.23 (0.24) [0.12]	-0.41 (0.52) [0.11]	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]		0.06 (0.5)	0.34 (0.55) [0.58]	0.13 (0.41) [0.44]	-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]	-0.23** (0.1)	-0.31 (0.2) [0.08]	-0.24** (0.11) [0.16]	-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]	-0.11 (0.22)	-0.19 (0.22) [0.04]	-0.09 (0.19) [0.04]	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]		-0.0324	-0.35 (0.3) [0.05]	-0.27** (0.12) [0.06]	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]	0.43 (0.54)	1 (0.95) [2.42]	0.44 (0.51) [0.69]	-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]	-0.09 (0.37)	0.03 (0.27) [0.32]	-0.05 (0.32) [0.08]	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]	0.07 (0.23)	0.51* (0.21) [0.08]	0.11 (0.22) [0.08]	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]	-0.22 (0.85)	0.42 (0.69) [1.61]	0.11 (0.72) [0.91]	-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]	-0.31** (0.15)	-0.22 (0.29) [0.11]	-0.38** (0.15) [0.13]	-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]	0.06 (0.1)	0.14 (0.18) [0.06]	0.07 (0.09) [0.07]	-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]	-0.5 (0.64)	-0.39* (0.33) [0.04]	-0.45 (0.55) [0.07]	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]	0.66* (0.36)	0.19 (0.87) [0.26]	0.54 (0.33) [0.69]	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]	-0.13 (0.21)	-0.06 (0.32) [0.03]	-0.11 (0.19) [0.04]	-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]	0.67 (0.88)	-2.44 (1.69) [1.55]	0.51 (0.85) [1.07]	-0.45 (0.47)	-1.23 (0.97) [1]	-0.57 (0.47) [0.82]
All other offenses except traffic				0.1 (0.11)	0.06 (0.12) [0.16]	0.1 (0.12) [0.08]	-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]	-0.36 (0.64)	-0.53 (245.1) [0.24]	-0.45 (0.56) [0.94]	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]	-0.66 (0.79)	-0.58***(0.19) [0.28]	-0.55 (0.71) [0.27]	-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]	-0.39 (0.68)	-0.2 (0.27) [0.18]	-0.33 (0.63) [0.17]	-0.44** (0.2)	-0.32 (0.32) [0.1]	-0.45** (0.19) [0.14]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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 DID column estimates are 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-selection post-lasso on the list of contemporaneous controls.

Table A5: 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]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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 DID column estimates are 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-selection post-lasso on the list of contemporaneous controls.

Table A6: 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	-0.144	-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]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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 DID column estimates are 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-selection post-lasso on the list of contemporaneous controls.

Table A7: Victimization cost of shale boom

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]

Notes: Enclosed in the [ ] reports mean square percentage error and ( ) reports a standard error. The 1%, 5%, and 10% levels of significance are given as ***, **, and *, respectively. GSC's standard errors 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 DID column estimates are 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-selection post-lasso on the list of contemporaneous controls.