Appalachian State University

Department of Economics Working Paper

Number 21-12| October 2021

Smoke and Fears: The Effects of Marijuana Prohibition on Crime

Scott Callahan
USDA Economic Research Service

David M. Bruner
Appalachian State University

Chris Giguere

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Smoke and Fears: The Effects of Marijuana Prohibition on Crime

Scott Callahan[1], David M. Bruner[2] and Chris Giguere[3]

Abstract

U.S. drug policy presumes prohibition reduces crime. Recently states have enacted medical marijuana laws creating a natural experiment to test this hypothesis but is impeded by severe measurement error with available data. We develop a novel imputation procedure to reduce measurement error bias and estimate significant reductions in violent and property crime rates, with heterogeneous effects across and within states and types of crime, contradicting drug prohibition policy. We demonstrate uncorrected measurement error or assuming homogeneous policy effects leads to underestimation of crime reduction from ending marijuana prohibition.

Keywords: Prohibition, Medical Marijuana Laws, Uniform Crime Report, Multiple Imputation

JEL Classifications: K42, C81

Acknowledgements

We would like to thank Todd Cherry, Bruce Benson, and seminar participants at Appalachian State University for their helpful comments. Any remaining errors are the sole responsibility of the authors. This research was supported in part by the U.S. Department of Agriculture, Economic Research Service.

Disclaimer

The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

[1] Corresponding Author: USDA Economic Research Service, Kansas City, Missouri, 64030 Email: Scott.Callahan@usda.gov

[2] Professor of Economics, Appalachian State University, Boone, North Carolina, 28608 Email: brunerdm@appstate.edu.

[3] Unaffiliated, Washington D.C., 20001 Email: giguerecs@gmail.com

Introduction

Drug policy in the U.S. is predicated on the notion that there is a causal link between crime and drug usage. Drug policy makers have stated, "Efforts to reduce the supply of drugs and enforce the laws of the U.S. are focused on decreasing crime."[4] Accordingly, U.S. drug policy has been largely focused on prohibition in an effort to stem the supply of drugs. As a result, the number of prisoners incarcerated for drug-related offenses rose 10-fold between 1980 and 2000, far outpacing increases in drug-related arrests which more than tripled (Kuziemko and Levitt, 2004). Still, it is not obvious that prohibition should reduce crime since much of the crime associated with illegal drugs, such as turf wars, punishment and retaliation, robbery, and theft, are due to the illegality of the product itself (Goldstein and Brownstein, 1987; Resignato, 2000). That is, in illegal markets where there is an absence of government provided property rights, market participants must protect themselves from predation and enforce contracts through the threat and use of violence (Rasmussen and Benson, 1994; Levitt and Venkatesh, 2000). Moreover, enforcement of drug prohibition can divert scarce resources away from the deterrence of other types of crime (Benson and Rasmussen, 1991; Benson et al., 1992). For instance, it is estimated marijuana legalization alone could save nearly $8 billion annually in averted enforcement costs (Pearl, 2018). Moreover, legalization creates a significant source of tax revenue that can be spent on deterrence, or invested in human capital, to raise the opportunity cost of crime (Lochner, 2004).[5] These arguments demonstrate there are sound reasons to question whether drug prohibition policies reduce crime.[6]

In recent years a handful of studies have focused specifically on the link between marijuana prohibition and crime (Alford, 2014; Morris et al., 2014; Huber et al., 2016; Gavrilova et al., 2017; Chu and Townsend, 2019). These studies exploit the growing trend of medical marijuana laws (MMLs) in the U.S., where individual states have repudiated federal laws to permit the production and consumption of marijuana for medicinal purposes, to test whether prohibition reduces crime. If prohibition is an effective deterrent, then we should observe an increase in crime in states that effectively end prohibition by passing MMLs, all else equal.[7] Yet, even though this natural experiment appears to be well-suited to testing this hypothesis, there remain significant challenges to obtaining valid estimates of the policy effect.

Notably, each of these studies of the effect of MMLs on crime has used the Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) program data. Similar approaches to policy analysis using the FBI UCR data have been conducted on topics such as gun laws (Lott and Mustard, 1997; Bronars and Lott, 1998; Dezhbakhsh and Rubin, 1998; Duggan, 2001), abortion laws (Donohue III and Levitt, 2001), labor market conditions (Raphael and Winter-Ebmer, 2001; Gould et al., 2002), environmental policy and resource discovery (Reyes, 2007; Liao et al., 2015; James and Smith, 2017), and crime deterrence (Chalfin and McCrary, 2017). Despite the popularity of the UCR data, it has severe deficiencies that make its use for policy analysis problematic (Maltz and Targonski, 2002). The program requests every police agency in the U.S. voluntarily report criminal offenses on a monthly basis to the FBI. Since there are

[4] See page 4 of the "National Drug Control Strategy" published by the Executive Office of the President of the United States in 2016.

[5] Washington state collected $923 million in excise taxes on marijuana retail from 2014 to 2018 (Hansen et al., 2017).

[6] Dills et al. (2008) provides evidence that among all the factors that economists believe to be determinants of crime the only factor that appears to be robustly correlated with crime is prohibition, which has a positive correlation with crime over a long period of time and across countries. Miron (1999) provides evidence that suggests prohibition leads to increases in homicides.

[7] Chu (2014); Wen et al. (2015) note the rise in marijuana usage in states that enact MMLs is evidence these laws essentially end prohibition.

no consequences for noncompliance, however, a significant number of reports are not submitted. To make matters worse, the FBI does not distinguish these missing values from true zeros in the UCR data. Hence, the missing data problem is severe and has significant potential to influence the inferences drawn from any policy analysis that uses this data.[8]

In this paper, not only do we improve upon previous estimates of the effect of MMLs on crime, but importantly, in doing so employ statistical techniques to minimize measurement error bias that can be applied to any policy analysis on crime in the U.S. Specifically, we use a multiple imputation procedure for agency-level crime data to fill in the gaps in the UCR data that accounts for the inherent uncertainty in these imputed values in the subsequent statistical analysis. This imputation procedure significantly improves upon prevalent imputation procedures, which have long been known to be insufficient (Maltz and Targonski, 2002). We then apply these imputed data to test our theoretically derived predictions, which include previously unexplored heterogeneous effects of MMLs on crime based urban density. This combination of improved data and model specification allows us gain insights into the true effect of MMLs on crime that are not readily apparent from previous studies.

Our results indicate that MMLs result in significant reductions in both violent and property crime rates, with larger effects in Mexican border states. While these results for violent crime rates are consistent with previously reported evidence (Gavrilova et al., 2017), we are the first paper to report such an effect on property crime as well.[9] Moreover, the estimated effects of MMLs on property crime rates are substantially larger, which is not surprising given property crimes are more prevalent. We also find novel evidence consistent with our hypothesis that MMLs reduce violent crime rates more in urban counties compared to rural counties, contrary to previous estimates (Chu and Townsend, 2019). We attribute this result to greater conflict between producers in urban counties under prohibition. Overall, our results are consistent with the need for market participants to create de facto property rights under prohibition, often through the use of violence. Our results are also consistent with prohibition causing a diversion of scarce policing resources, which when reallocated have the greatest impact on more pervasive types of crime and in locations where crime rates are higher. These findings demonstrate both the importance of accounting for heterogeneous policy effects on crime and the necessity to correct for measurement error in crime data when conducting policy analysis.

The remainder of the paper is organized as follows. In section 3, we review the history of MMLs in the U.S. and related literature relevant to the present study. In section 5, we present our imputation approach and describe our empirical model. In section 6, we present our main results and the results of robustness checks. Lastly, summarize our findings and discuss their implications in section 8.

Theory

We present a simple model of conflict over rents in an illegal market with insecure property rights to motivate our empirical analysis (Hirshleifer, 1995; Grossman and Kim, 1995). Suppose there are $N \geq 2$ identical players, indexed $i = 1, \ldots, N$. Each player has an endowment of ω of resources. Agents can allocate resources to an investment, k_i, with an exogenous rate of return, A, that is common to all agents. However, production, Ak_i, is insecure since property rights are not well-defined. Hence,

---

[8] Maltz and Targonski (2002) attribute the widely criticized conclusion of Lott and Mustard (1997) that “right-to-carry” gun laws reduced homicide and other violent crime rates to the use of UCR data aggregated to the county-level.

[9] Gavrilova et al. (2017) find no evidence that MMLs significantly reduce violent crime rates in states that do not share a border with Mexico.

resources can also be devoted to conflict, x_i, in order to secure production.[10] Each player simultaneously maximizes his payoff by choosing how he allocates his endowment, ω, across the two possible choices: production and conflict:

$$\omega = k_i + \theta x_i, \qquad \forall\, i = 1, \dots, N, \tag{1}$$

where θ denotes the relative price of conflict. Let p_i denote the proportion of production i appropriates given by

$$p_i = \frac{x_i}{X}, \qquad \forall\, i = 1, \dots, N, \tag{2}$$

where $X \equiv \sum_{i=1}^{N} x_i$. Thus, $p_i + \sum_{j \neq i}^{N} p_j = 1$ for all $i = 1, \dots, N$. We assume each player's payoff is the amount production he appropriates:

$$u_i = p_i A\left(k_i + \sum_{j \neq i}^{N} k_j\right), \qquad \forall\, i = 1, \dots, N. \tag{3}$$

An agent's optimal allocation is determined by substituting equations (1) and (2) into (3) and equating the derivative with respect to x_i to zero,

$$\frac{\partial u_i}{\partial x_i} = \frac{X_{-i}}{(X)^2} A\left(\omega - x_i + \sum_{j \neq i}^{N} (\omega - \mathrm{x_j})\right) - A p_i = 0, \qquad \forall\, i = 1, \dots, N. \tag{4}$$

Thus, agents tradeoff a larger pie for a larger slice of the pie at the margin. By imposing symmetry, the Nash equilibrium solutions are:

$$x_i{}^* = \frac{\omega(N-1)}{N}, k_i{}^* = \frac{\omega}{N-1}, \qquad \forall\, i = 1, \dots, N. \tag{5}$$

Hence, in equilibrium, investment and conflict are complimentary; agents do not engage in one without the other. Moreover, conflict is increasing in the size of the endowment and the number of agents,

$$\frac{\partial x_i^*}{\partial \omega} = \frac{N-1}{N} > 0, \qquad \frac{\partial x_i^*}{N} = \frac{\omega}{N^2} > 0, \quad \forall\, i = 1, \dots, N. \tag{6}$$

By contrast, when property rights are well defined the proportion of production, AK, i appropriates is proportional to i's production, Ak_i,

$$p_i = \frac{k_i}{K}, \qquad \forall\, i = 1, \dots, N. \tag{7}$$

Since production is secure agents have no incentive to engage in conflict,

$$x_i{}^* = 0, k_i{}^* = \omega, \quad \forall\, i = 1, \dots, N. \tag{8}$$

We use this simple theoretical framework to motivate testable hypotheses for our empirical investigation of the effects of MMLs on crime.

Hypothesis 1 *Equation (5) implies producers in an illegal marijuana market will engage in conflict to appropriate insecure production. Since MMLs legalize the marijuana market and create enforceable*

[10] We assume A is large enough the optimal solution is to allocate one's entire endowment between production and conflict.

property rights, equation (8) implies we should observe reduced crime due to a decrease in conflict between marijuana producers.

Our next hypotheses relate to expected differential effects of MMLs on crime. Firstly, we hypothesize a differential impact based on urban density.

Hypothesis 2 *Assuming the number producers operating in illegal marijuana markets is proportional to the population density, equation (6) implies crime associated conflict over the insecure production should be higher in urban counties compared to rural counties. Hence, MMLs should reduce crime more in urban counties than rural counties.*

Next, Gavrilova et al. (2017) note that drug trafficking organizations (DTOs) operate primarily in states that border Mexico (Finklea et al., 2010). These organizations make substantial profits with an estimated $1.5 billion a year due to marijuana exports alone (Kilmer et al., 2010).

Hypothesis 3 *Since DTOs operate primarily in Mexican border states and have large sums of money at their disposal, equation (6) implies crime associated conflict over the insecure production should be higher in border states compared to non-border states. Thus, MMLs should reduce crime more in border states than in non-border states.*

Lastly, we expect MMLs to have a differential effect on violent and property crime due to the reallocation of scarce policing resources. Since more severe crimes carry harsher penalties to provide greater deterrence (Becker, 1968), violent crimes are less prevalent than property crimes. Our prediction relies on the assumption that the marginal product of policing effort (i.e., the marginal reduction in crime) is higher when crime is high; crime is decreasing in policing effort at a decreasing rate. We also assume that the reallocation of scarce policing resources is proportional to the amount of conflict under prohibition.

Hypothesis 4 *If crime is decreasing in policing effort at a decreasing rate, MMLs should reduce property crime more than violent crime due to the reallocation of scarce policing resources since property crime is more prevalent. Further, there should be a greater reallocation of policing resources when conflict is higher under prohibition, namely in border states and urban counties. So MMLs should reduce property crimes more in border states and urban counties.*

Background

Medical Marijuana Laws and Crime

According to the World Health Organization, cannabis is the most widely used drug in the world, with estimates of roughly 3.9 percent of the global population, and 8.0 percent of the U.S. population, between the ages of 15 and 64 years having used it at least once in 2016 (United and Nations, 2018). Cannabis was widely used for medicinal purposes in the U.S. during the 19th and early 20th centuries (Holland, 2010), until passage of the Marihuana Tax Act in 1937 (Musto, 1972). Strict prohibition under federal law began in 1970 with passage of the Controlled Substances Act, where the Federal Drug Enforcement Agency defined it as a Schedule 1 substance, i.e., it has no accepted medical use and high risk of addiction (Carliner et al., 2017). The Schedule I classification has made medical research on the efficacy of cannabis extremely difficult, yet therapeutic agents based on tetrahydrocannabinol (THC), the major psychoactive component of cannabis, such as dronabinol have been approved for use as an antiemetic for years (Bridgeman and Abazia, 2017). California was the first state to enact a MML in 1996 (Mark Anderson et al., 2013). As of 2016, thirty-three states and the District of Columbia permit the sale of medicinal marijuana (see table 1). The most common medical conditions accepted by states for marijuana use are related to the relief of symptoms associated with cancer, glaucoma, HIV/AIDS, and MS (Bridgeman and Abazia, 2017). Yet, despite its growing acceptance for medicinal usage, the question of whether relaxing prohibition on marijuana will result in an increase in crime remains unclear.

Previous studies that have examined MMLs effects on crime are summarized in Table 2. Despite using the same data and the same difference-in-differences statistical technique, these studies have arrived at different conclusions. Given differences in the time periods analyzed, geographic jurisdiction considered (i.e., level of aggregation), types of crime considered, and empirical model specifications, it is not surprising that results vary across studies. For example, Alford (2014) only examined property crimes and murder, but not other violent crimes like aggravated assault. Her analysis allowed for the differential effects of MMLs based on attributes such as the existence of dispensaries, which had a positive correlation with robbery rates, and the ability for home cultivation, which was negatively correlated with robbery rates.[11] Morris et al. (2014) and Huber et al. (2016) conducted similar state-level analyses on individual types of violent and property crime, but used different specifications of the effect of MMLs on crime and looked at different time periods. While Morris et al. (2014) estimated MMLs have practically no effect on crime rates at the state- level, Huber et al. (2016) estimated MMLs significantly reduce robbery, burglary, and larceny rates. Gavrilova et al. (2017) conducted a county-level analysis on violent crime rates allowing MMLs to have differential effects in Mexican border states.[12] Their analysis suggested MMLs were negatively correlated with violent crime rates in states that shared a border with Mexico, with the effect of MMLs on violent crime rates decreasing with distance to the border. Further evidence suggests this effect was due to crimes associated with drug trafficking organizations (DTOs). Meanwhile, Chu and Townsend (2019) conducted an agency-level analysis of agencies in cities with more than 50,000 residents. They focus their analysis on large cities, which are more likely to report to the FBI, to avoid endogenous sample selection (Akiyama and Propheter, 2005; Lynch and Jarvis, 2008). They estimated models with linear, quadratic, and cubic city-specific time trends and found no robust evidence that MMLs are correlated with crime rates.

These studies attempted different strategies to minimize measurement error bias due to the missing data problem endemic in the UCR data. However, given these strategies all try to avoid the problem, their ability to accurately estimate the effect of MMLs on crime is unlikely. State- level analyses have the benefit of minimizing the impact of missing data, but cannot estimate heterogeneous effects of MMLs on crime within a state. Studies that have used less aggregated UCR data to obtain sufficient statistical power attempt to minimize measurement error bias by restricting their samples to urban areas to avoid missing data.[13] However, this approach omits the vast majority of the country from the analysis, severely limiting the relevant population for which conclusions can be drawn. We adopt a different approach that

---

[11] Burkhardt and Goemans (2019) also examined the impact of dispensaries on crime, but at a smaller scale. Importantly, they demonstrated heterogeneous effects. In particular, they found that dispensary openings in Denver, Colorado were associated with "decreases [in] violent crime rates in above median income neighborhoods" (Burkhardt and Goemans, 2019, p.). Furthermore, "non-marijuana drug-related crimes decrease within a half-mile of new dispensaries but do not simultaneously increase within a half-mile to mile of new dispensaries" (Burkhardt and Goemans, 2019, p.). One the other hand, Markowitz (2005) found that decriminalization was associated with a higher probability of assault, using a nationally representative sample from the 1990's

[12]The county-level data on violent crime rates were collected from the NACJD, and are based, in part, on naive imputation methods that fail to account for the uncertainty in such guesses. We discuss this in detail in the next section.

[13] Gavrilova et al. (2017) use naively imputed values that are not distinguished from observed values in their main analysis and perform robustness checks that eliminate any counties that have any imputed values.

employs a multiple imputation procedure to fill in missing data and accounts for the inherent uncertainty of the imputed values in the statistical analysis thereby reducing measurement error bias. This approach enables us to use less aggregated data and estimate heterogeneous effects of MMLs on crime within a state, while maintaining a sample representative of most of the country, to obtain a more accurate estimation of the effect of marijuana prohibition on crime.

FBI Uniform Crime Reporting

The FBI began to compile data to document "crime in the United States" in the 1930's. At present the UCR data are composed of seven different data sets: (i) offenses known to the police (Return A), (ii) age, sex, race, and ethnicity of arrestees (ASR), (iii) law enforcement officers killed or assaulted (LEOKA), (iv) police employment, (v) arson reports, (vi) supplemental homicide report (SHR), and (vii) the hate crime supplement (Lynch and Jarvis, 2008). Over the years the program has expanded in a number of ways. For example, the National Incident-Based Reporting System (NIBRS) was created to improve the quality of these data (Liao et al., 2015). There have also been substantive changes made to the data. For example, the Violent Crime Control and Law Enforcement Act (VCCLEA) of 1994 banned assault weapons, and defined hate, sex, and gang- related crimes and thus impacted how crime was reported. Similarly, the definition of rape was changed in 2013.[14]

As previously mentioned, the voluntary nature of the program creates a missing data problem that adds additional complexity to the use of UCR data for policy analysis. Crimes are reported by individual police agencies, which is compiled and aggregated to both the county- and state-level. Aggregated data are available from two sources, the FBI and the National Archive of Criminal Justice Data (NACJD). Both the FBI and NACJD impute missing values prior to aggregation. Boylan (2019) characterizes three broad imputation methods used in the UCR data. First, the FBI creates the state-level Return A data by replacing missing data with average crime rates from comparable agencies in a given state-year (Targonski, 2011). The NACJD has adopted a similar procedure for their county-level data since 1994.[15] Before 1994 the NACJD replaced missing values with average crime rates from a given county-year (Boylan, 2019). Finally, some scholars have used "longitudinal" methods such as replacing missing data with past values from the same agencies. Each of these naive methods, however, ignore the uncertainty surrounding the imputed values which mitigates their use in policy analysis. Maltz and Targonski (2002), Lynch and Jarvis (2008), and Targonski (2011) have all called for the development and use of improved imputation methods for the UCR data.[16]

---

[14] Specifically, the term "forcible" was removed from the offense name, and the definition was changed to "penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim." (U.S. Federal Bureau of Investigation (FBI), 2020).

[15] There are two main components of their imputation procedure applied to UCR data since 1994. First, if an agency does not report a full twelve months of data, but has reported at least three months of data, NACJD inflates this number to estimate the annual crime total. Second, for agencies that have less than three months of reporting, an value is imputed by the average of crimes committed in the same state, year and population bin as the agency with the missing value. The county-level data published by the FBI implicitly assumes that missing agency-level observations are zero prior to aggregating (Targonski, 2011). In addition, the NACJD and FBI data differ for two other reasons. First, the NACJD allocates crimes reported by state agencies across counties by population (Maltz and Targonski, 2002; National Association of Criminal Justice Data (NACJD), 2016). Second, the NACJD corrects problems with the FBI's population data that leads to double-counting (Maltz and Targonski, 2002).

[16] Targonski (2011) re-examined the UCR data and identified the nature and extent of the missing data problem, developed methods to clean the data, and tested the FBI procedure against an alternative. Both Liao et al. (2015) and Boylan (2019) have built on the this work. Boylan (2019) concludes that smaller law enforcement agencies are

Data

We focus on the Return A, or Table 1, data in the FBI UCR in the present study.[17] Our primary variables of interest are violent and property crime rates, which can have differential responses to policy change (Cherry and List, 2002). We independently define violent crimes as murder, robbery, and aggravated assault; rape is excluded because of the change in definition to prevent any potential confounding effect. Likewise, we define property crime as burglary, larceny, and motor vehicle theft; arson is excluded because we seek to focus exclusively on stealing rather than destruction of property. Consequently, our measures of violent and property crime may be different from those found in published UCR sources but result in measures that are consistent over time. Our time series begins in 1994 to coincide with the passage of the VCCLEA, which represents a major shift in federal crime policy, to prevent any confounding effects associated with this legislation. We use data only for city police agencies and county sheriff departments, which we aggregate by county.[18] County aggregate socio-economic control variables, including population size, the percentage of the population that is male, age 10 to 19, age 20 to 24, black, and Hispanic are obtained from the U.S. Census Bureau's intercensal estimates for the years 1994 to 2010 and from the American Community Survey from 2010 to 2016.[19] Estimates of the county unemployment rate, collected from the U.S. Bureau of Labor Statistics, are used to control for labor market conditions associated with the opportunity cost of crime. We construct an indicator variable to distinguish urban and rural counties. Specifically, an urban county is defined as one that contains a city with a population of more than 50,000.[20] Finally, we also construct an indicator variable to distinguish counties in states that share a border with Mexico from those that do not.

Imputation Procedure

We utilize multiple imputation to correct for measurement error in agency-month crime rates. The goal is to estimate a county-level crime rate which can be directly compared with rates at other points in time and geographical location. While the county aggregate from this approach is not perfectly representative of the true crime rate, the estimates resulting from the imputation can be directly compared with each other to obtain valid estimates of policy effects in a panel data framework.

The first step towards justifying the use of any imputation procedure is to characterize the nature of the missing data mechanism. There are three types of missing data mechanisms: (i) missing completely at random (MCAR), (ii) missing at random (MAR) and, (iii) missing not at random (MNAR). If data are MCAR, then empirical analysis of the non-missing data is consistent and the missing data problem can be ignored. Unfortunately, since crime rates are more likely to be missing from the UCR for smaller agencies and jurisdictions, the missing data problem cannot be ignored. If data are MNAR, then the propensity for

---

less likely to report to the FBI when crime is high, which causes an attenuation bias and thus leads to underestimates of policy effects.

[17] Table 1 of the FBI's annual report present data on murder, rape, robbery, aggravated assault, burglary, larceny, motor vehicle theft, and arson.

[18] Statewide law enforcement agencies enforce laws state-wide, rather than in specific local jurisdictions. As such, to include these agencies in our analysis would require an assumption about the spatial distribution of their activities. NACJD assigns crimes reported by state agencies to counties according to the population weight of the county. However, it is likely state agencies are more active in law enforcement in rural areas than in urban areas. However, it is empirically impossible to measure the degree to which this is true. Since we seek imputed values with internally consistent variation, rather than accurate level estimates, we choose to omit statewide agencies and any other agency reporting zero population from the analysis.

[19] The intercensal contain annual interpolated county estimates based on the 1990, 2000 and 2010 decennial censuses.

[20] We only include cities where law are enforced by police agencies, rather than sheriff's departments.

an observation to be missing is directly dependent on the value of the missing observation, making imputation practically impossible. A missing data mechanism is MAR if the value of a missing observation is conditional on non-missing variables.[21]

We assume that the missing data mechanism (ignoring for the moment the nature of true zeros versus missing values) is missing at random, such that the value of the missing crime rate depends not on the crime rate itself, but on the size of the police agency. Our imputation strategy relies on assuming that the size of the agency is proportional to the population of the jurisdiction. Since Boylan (2019) suggests smaller agencies do not report when crime is high, we conduct robustness checks that inflate imputed values by 25% and 50% to verify the sensitivity of our results to this assumption.

Multiple imputation is a variant of stochastic imputation. Stochastic imputation adds a stochastic element to the imputed values, in order to construct estimates that reflect the degree of uncertainty about the accuracy of the imputed data. Multiple imputation uses multiple stochastic imputations in order to calculate standard errors which accurately reflect the uncertainty about the parameter estimates obtained using imputed data. Typically, multiple imputation is a three step process. First, *n* imputations are conducted in order to obtain n imputed datasets. Second, *n* estimations are conducted. Third, the coefficients and standard errors from the *n* estimations are combined to obtain valid coefficients and standard errors.

Since the missing data mechanism is at the agency-month level, and the analysis is at the county-year, we modify this procedure. First, we obtain 100 imputed datasets using multiple imputation. After removal of extreme values and agencies which cannot be imputed reliably, a process which will be described later on, we then aggregate the agency-month data to the agency- year level.[22] Then we aggregate the agency-year level data to the county-year level. We do so because we were unable to find municipality level control variables prior to the introduction of the American Community Survey.[23] We then conduct 100 estimations, one for each dataset. Finally, we combine the results to obtain unbiased coefficients and corrected standard errors.

The first step of our procedure is conducting agency-year level imputations. We use the fully conditional specification to conduct the imputations[24]. We conduct separate imputations for each state because crime is a function of latent characteristics such as local culture, law enforcement policy and legal regimes that vary by state. We then model the agency-month crime rates with the following model.

$$y_{asmy} = \alpha + \gamma_a + \delta_t + \theta_m + \beta\, log(POP_a st) + \varepsilon_a st \tag{9}$$

In the above empirical model, y_{asmy} is the crime rate, α is the intercept, γ_a represents an agency fixed effect, δ_t represents the year fixed effect, θ_m denotes the month fixed effect[25], β is the coefficient on the log of agency population, which is denoted by POP_{ast}, and ε_{ast} represents the error term.[26] In cases where there are too few observations to calculate this agency fixed effect, extreme outliers result. For

[21] See Seaman et al. (2013) for a detailed explanation of missing data assumptions.
[22] To the best of our knowledge, we are the first to introduce an aggregation step between the imputation and estimation steps.
[23] We wish to note, however, that for studies focused on a more recent time series, an agency-level analysis would be preferable, and this procedure can easily accommodate this..
[24] See van Buuren (2007) for more details.
[25] The monthly data identify the month of each observation, allowing us to use month fixed effects to control for possible seasonality in the crime rate data.
[26] These imputations are conducted using MI procedure in SAS.

this reason, we use two methods to remove extreme outliers. First, crime rates that are imputed to be less than zero are set equal to zero. Second, we delete the agency from the entire analysis if any imputed crime rate exceeds three standard deviations above the mean for a given year.[27]

Next, we aggregate the agency-month data to the agency-year level. We do so by converting crime rates back to crime totals and summing crime to the agency-year level, and setting aside annual agency population totals. Then, we aggregate crime and population totals to the county-level. We aggregate each imputed data-set separately by summing crime totals and population totals at the county-level, and then calculating the crime rate. The resulting county aggregate crime rates are relatively consistent at different points of time and geographical locations. Once aggregated to the county-level, we merge these data with county-level demographic control variables. We construct our policy variables of interest using information on the implementation of MMLs, urban classification, and status as a Mexican border state.

Lastly, we conduct 100 panel estimations per empirical model, one for each imputed dataset, and combine these results.[28] This corrects standard errors to reflect the degree of uncertainty about the accuracy of the imputed values. This process is repeated for each regression specification and type of crime. Fit statistics are averaged over the estimations for each specification.

True Zeros vs. Missing Observations

The second major problem with the UCR data is that no distinction is made between missing values and true zeros. This is serious problem for smaller jurisdictions, which constitute the majority of jurisdictions in the U.S. Murder is a particularly rare crime. It is the norm rather than the exception that rural areas with small populations rarely report any murders. We posit that it is most reasonable to assume that if any crime category is reported by an agency, the remaining blanks are zero. While this assumption is not perfectly accurate, we believe it to be the most sensible assumption to make.

Consider the pattern of missing data by crime reported in tables 3 and 4. First, 35.8 percent of agency-month observations are missing all crime rates, which demonstrates the necessity of using imputation to fill in the gaps. 10.6 percent of the observations are missing murder rates, with all other crimes reported. Agencies reporting all crime rates but murder and robbery constitute 9.4 percent of observations. Only 2.4 percent of observations report values for all crime rates. There is a clear pattern. When crime data are reported, data are more likely to be missing for more serious or uncommon crimes. We use this evidence to justify using our assumption. To do otherwise will omit the vast majority of low population density jurisdictions from the analysis. The results of analyses of such data would be unrepresentative of the many rural communities in the U.S.

To demonstrate the importance of the missing data problem in the UCR and the pitfalls of the NACJD imputation method outlined in Maltz and Targonski (2002), we include an analysis using NACJD data as a control group to assess how our imputation procedure affects empirical results.[29] This estimation

---

[27] We avoid a fundamental problem, the one we are attempting to correct for, by removing the entire agency from the analysis. Specifically, we avoid having to address the question of whether county-level crime rates could be changing due to omitted agency-level data or changes in true crime rates. We do not claim that this imputed data perfectly represents the true crime rate. Rather, this procedure allows for relative consistency of estimated policy effects by correcting the omitted agency problem. Note that many agencies were created during the time series. Unfortunately, we are unaware of a data-set that tracks the dates these agencies were created. We therefore assume that the first year in which an agency observed in the UCR is the first year that the agency existed.

[28] This is done using the PANEL and MI ANALYZE procedures in SAS.

[29] Due to challenges obtaining an extended time series from NACJD, we instead use violent crime data provided by Gavrilova et al. (2017), which they obtained from NACJD, for our control group

replicates the estimations conducted by Gavrilova et al. (2017), though we use our empirical model in order to better compare results between the dataset provided by NACJD and our imputed dataset.

Summary Statistics

Table 5 reports sample sizes for the data set imputed using multiple imputation. The end result is a data set that preserves the data of 14,601 to 16,276 police and sheriff's departments. From this, we build a data set with excellent county coverage. Figure 1 shows the geographic distribution of missing crime data. Our procedure successfully imputes data for the vast majority of counties. Only a handful of the most rural counties in the U.S. are omitted.

Summary statistics for model variables are presented in table 6. Of note are the summary statistics for imputed crime rates. The means (taken over all of the imputations) of the violent and property crime rate variables are comparable to official estimates. When comparing the NACJD data obtained from Gavrilova et al. (2017) to our multiply imputed data over the same time series, sample sizes for our imputed data are slightly smaller, indicating a handful of county-year observations are dropped due to extreme outliers in the imputation step. Further, with our multiply imputed data, the mean and standard deviation of the violent crime rate is modestly higher than the NACJD data used by Gavrilova et al. (2017), but quite similar. While these similarities are a positive indication for success, we must reiterate that similarities to official estimates are neither the goal nor required; our goal is to preserve the relative temporal and cross-sectional variation in crime rates in order to conduct valid policy analysis.

Empirical Model

We estimate the crime rate y_{cst} in county c, in state s at time t as a function of an intercept term α_0, county fixed effects γ_c, time fixed effects δ_t, and state specific time trends, denoted by λ_{st}. We use these linear time trends to control for both the general decrease in crime rates observed nationally over the duration of this time series, and allow this trend to vary based on state law, culture, and other unobserved state specific characteristics. We also include previously stated county specific socioeconomic control variables X_{cst}, as well as an indicator variable MB_s to control for counties in Mexican border states, and an indicator variable $URBAN_c$ to identify counties with cities that have populations in excess of 50 thousand people.[30] We estimate the effect of MMLs on crime by including the variable MML_{st}, which is a variable indicating that medical marijuana laws have been passed in state s at time t, which we interact with MB_s and $URBAN_c$, to allow the effect of MMLs to differ between border and non-border states and between urban and rural counties, respectively. Finally, we include an interaction between $MML_s t$, MB_s, and $URBAN_c$, to allow for the any difference in the effect of MMLs between urban and rural counties to differ between border and non-border states.

Hence, our main empirical specifications for both violent and property crime rates is given by:

$$y_{cst} = \alpha_0 + \gamma_c + \delta_t + \lambda_{st}t + \theta X_{cst} + \beta_1(MML_{st} \times URBAN_c) + \beta_2(MML_{st} \times URBAN_c) \\ +\beta_3(MML_{st} \times MB_s) + \beta_{\ 4(MML_{st} \times MB_s \times URBAN_c)} + \varepsilon_{cst} \tag{10}$$

To demonstrate the importance of controlling for heterogeneous effects of MMLs on crime, we include estimations that impose restrictions on this specification. Namely, model 1 restricts $\beta 2 = \beta 3 = \beta 4 = 0$, model 2 restricts $\beta 2 = \beta 4 = 0$, model 3 restricts $\beta 3 = \beta 4 = 0$, while model 4 restricts $\beta 4 = 0$. Model 5 relaxes all restrictions and corresponds to the full specification in equation (10).

[30] We have suppressed these variables in the equation (10) and the results tables for the sake of brevity.

Results

Our main estimation results for violent crime rates and property crime rates are reported at the top and bottom of table 7, respectively. As can be seen, the estimated effects of MMLs on crime rates are fairly robust to the previously stated restrictions, but the restricted models tend to underestimate to total effect of MMLs on crime rates, particularly for violent crime. This overall result demonstrates the importance of accounting for potential heterogeneous effects in policy analysis on crime. Accordingly, we will focus our discussion of the results on fifth column, corresponding to the unrestricted model specification in equation (10).

We will first summarize the results for violent crime rates. The estimation results indicate an average reduction in the violent crime rate of a county of roughly 29 violent crimes annually in response to MML implementation. In counties of states that share a border with Mexico, there is an average additional reduction of 61 violent crimes per year. In addition, the violent crime rate of urban counties is further reduced by 31 violent crimes per year on average in states with a MML. We find no significant difference in this effect between border and non-border states. Taken in total, these results suggest that not only does legalization of medical marijuana reduce violent crime rates, but that these violent crime rate reductions are larger for both urban areas and in Mexican border states.

Next we turn our attention to the estimated the effects of MMLs on property crime rates. The results indicate an average reduction of roughly 180 property crimes in a county per year in states with a MML. In addition, in counties of border states there is and additional reduction of 258 property crimes per year. We do not find robust evidence that urban counties experience a differential change in their property crime rates compared to rural counties in response to MMLs, regardless of whether they are in a border state or not. It is worth noting that the reductions in property crime rates are much larger than violent crime rates, which is not surprising since property crimes are far more common.

To demonstrate the measurement error bias that arises from using naively imputed values that are unaccounted for in the analysis, we report the results from a second set of estimations in table 8. We do this by using the same county-level NACJD violent crime rate data that Gavrilova et al. (2017) used to conduct their analysis. We constrain our time series to the same period (1994-2012) they used and omit Alaska and Hawaii as they did. We use our own set of control variables, rather than those they provided with their data.[31] The top of the table reports the results using our imputation procedure on their data and the bottom of the table reports the results using the NACJD imputed values.

Again, we will focus our discussion on the results reported in the fifth column of the unrestricted model. The results reported at the bottom of the table replicate the significant effect of MMLs on violent crime rates in counties of border states reported by Gavrilova et al. (2017).[32] By contrast, the results at the top of the table are nearly identical to our main estimation results reported in table 7. These results

[31] We initially attempted to conduct a longer time series for our test case, but had difficulty finding the appropriate data from NACJD. In particular, the data from 2015 were missing. In light of this, we chose to use the data obtained from Gavrilova et al. (2017) instead. We choose to use our own control variables for two reasons. One, we have omitted the poverty rate and median income, since we were unable to find these variables on the census website for the whole time series. Second, there was an error with the variable Gavrilova et al. (2017) used to control for the percentage of the black population; there were some observations in which the percentage exceeded 100 percent.

[32] We also estimate MMLs have a small and weakly significant negative effect on violent crime rates in counties of non-border states. We believe this is due to our use of standard errors that are clustered at the county level rather than at the state level.

demonstrate the attenuation bias caused by the missing data problem that leads to underestimates of policy effects (Boylan, 2019).

Robustness Checks

Given the novelty of this imputation procedure, we conduct three robustness checks designed to assess the validity of the assumption that crime rates are MAR, conditional on the fully observed jurisdiction population variable.

First, we conduct a simulation exercise where we create a single stochastic imputation of agency month crime rate. Then, we delete at random the same number of observations as are truly missing, by state. After conducting 100 imputations based on this simulated data set, we aggregate the data to the county level and perform an analysis using the same procedures as we use for our main results. If the data are truly MAR conditional on population, then the values of the latent and observed crime rates should not affect the statistical results, providing justification for the MAR assumption. If instead agencies are choosing not to report when their crime rates are abnormally high, then the results should differ from our main specification.

Table 9 contains the results of this simulated imputation procedure. Results for violent crime rate estimations are nearly identical, in both magnitude and statistical significance. The results for property crime, however, are not as consistent. While we still estimate a reduction in the property crime rate of counties in response to a MML, the magnitude is much larger. Furthermore, we find no evidence of a significant difference in the size of this effect between border and non-border states. This evidence suggests that the procedure works better for violent crime than for property crime.

For this reason, we conduct two additional robustness checks. Since Boylan (2019) suggests smaller agencies do not report when crime is high, Tables 10 and 11 report the results for estimations based on imputed crime rates inflated by 25 percent and 50 percent, respectively. In either case, the results for violent crime are nearly identical to our main results in Table 7. For property crime rates, when the imputed values are inflated by 25 percent the estimated effects of MMLs are of similar magnitude to our main results, but are substantially different when the imputed values are inflated by 50 percent. In either case, we continue to estimate a significant reduction in the property crime rates of counties in border states associated with MMLs. The primary takeaway from these last two robustness checks is that the estimated effects of MMLs on violent crime rates are quite robust, while we only find robust evidence of reductions in property crimes rates in Mexican border states.

Conclusion

In this paper we attempt to answer the question, does prohibition lead to a reduction in crime? To do so, we take advantage of the growing trend in medical marijuana laws in the U.S. that effectively end federal prohibition of marijuana. We employ a difference-in-differences empirical strategy to exploit natural variation in these laws across states and time. To estimate the effects of these policies accurately, however, we must first address the severe measurement error endemic in the FBI Uniform Crime Report data.

Therefore we develop a novel multiple imputation procedure to fill in the missing agency-level data. This procedure allows us to account for the inherent uncertainty associated with the imputed values when conducting our policy analysis on crime. This approach is the most rigorous attempt to data to minimize measurement error bias in any empirical study of crime in the U.S. using the UCR data and undoubtedly a significant improvement over commonly adopted practices. The implications of this contribution reach well beyond the present application, given the prominent use of the UCR data in the empirical literature.

Using this imputed data, we estimate significant reductions in violent crime rates in states that legalize medicinal marijuana. Moreover, we find evidence that ending marijuana prohibition results in larger reductions in violent crime rates in states that border Mexico and in urban counties. We also find evidence that medical marijuana legalization reduces property crimes, with larger reductions in states that border Mexico. Our results indicate that when this heterogeneity is not accounted for the total effect of medical marijuana legalization on crime is underestimated. We also demonstrate that previous estimates of the effect of medical marijuana laws on crime fail to appropriately address the problem of the measurement error in the crime data, have underestimated the reduction in crime from of ending marijuana prohibition due to attenuation bias.

Given the novelty of the imputation procedure, we perform several robustness checks to that explore the validity of our assumptions. The results for violent crime appear to be quite robust suggesting this procedure is well-suited to this data. While the results for property crime are less robust, we consistently estimate significant reductions in property crime rates associated with legalization of medical marijuana. One explanation for this discrepancy could be that the reductions in violent crime rates are primarily caused by increased property rights in the market while the reductions in property crime rates are mostly caused by a reallocation of policing resources. Thus, the effects of MMLs on property crime rates are more sensitive to population, and thus the imputed values, than violent crime rates. Perhaps future research can find better instruments or methods of imputing property crimes.

References

Akiyama, Y. and Propheter, S. K. (2005). Methods of Data Quality Control: For Uniform Crime Reporting Programs Cross-Sectional Outlier Detection Longitudinal Outlier Detection Proportional Outlier Detection. Technical report, Criminal Justice Information Services Division. *Federal Bureau of Investigation*.

Alford, C. (2014). How medical marijuana laws affect crime rates. *mimeo University of Virginia*.

Becker, G. S. (1968). Crime and punishment: An economic approach. In: *The economic dimensions of crime*, pages 13–68.

Benson, B. L., Kim, I., Rasmussen, D. W., and Zhehlke, T. W. (1992). Is property crime caused by drug use or by drug enforcement policy? *Applied Economics*, 24(7):679–692.

Benson, B. L. and Rasmussen, D. W. (1991). Relationship between illicit drug enforcement policy and property crimes. *Contemporary Economic Policy*, 9(4):106–115.

Boylan, R. T. (2019). Imputation methods make crime studies suspect: Detecting biases via regression discontinuity.

Bridgeman, M. B. and Abazia, D. T. (2017). Medicinal cannabis: history, pharmacology, and implications for the acute care setting. *Pharmacy and Therapeutics*, 42(3):180.

Bronars, S. G. and Lott, J. R. (1998). Criminal deterrence, geographic spillovers, and the right to carry concealed handguns. *The American Economic Review*, 88(2):475–479.

Burkhardt, J. and Goemans, C. (2019). The short-run effects of marijuana dispensary openings on local crime. *The Annals of Regional Science*, 63(1):163–189.

Carliner, H., Brown, Q. L., Sarvet, A. L., and Hasin, D. S. (2017). Cannabis use, attitudes, and legal status in the us: a review. Preventive medicine, 104:13–23.

Chalfin, A. and McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1):5–48.

Cherry, T. L. and List, J. A. (2002). Aggregation bias in the economic model of crime. *Economics Letters*, 75(1):81–86.

Chu, Y.-W. L. (2014). The effects of medical marijuana laws on illegal marijuana use. *Journal of Health Economics*, 38:43–61.

Chu, Y.-W. L. and Townsend, W. (2019). Joint culpability: The effects of medical marijuana laws on crime. *Journal of Economic Behavior & Organization*, 159:502–525.

Dezhbakhsh, H. and Rubin, P. H. (1998). Lives saved or lives lost? The effects of concealed-handgun laws on crime. *The American Economic Review*, 88(2):468–474.

Dills, A. K., Miron, J. A., and Summers, G. (2008). What do economists know about crime? Technical report, National Bureau of Economic Research.

Donohue III, J. J. and Levitt, S. D. (2001). The impact of legalized abortion on crime. *The Quarterly Journal of Economics*, 116(2):379–420.

Duggan, M. (2001). More guns, more crime. *Journal of Political Economy*, 109(5):1086–1114.

Finklea, K. M., Lake, J. E., Franco, C., Haddal, C. C., Krouse, W. J., and Randol, M. A. (2010). Southwest border violence: Issues in identifying and measuring spillover violence. Technical report, CRS Report for Congress R41075, Washington, D.C.

Gavrilova, E., Kamada, T., and Zoutman, F. (2017). Is legal pot crippling Mexican drug trafficking organizations? The effect of medical marijuana laws on us crime. *Economic Journal*, (239120):1–33.

Goldstein, P. J. and Brownstein, H. (1987). Drug related crime analysis-homicide. Narcotic & Drug Research Incorporated New York.

Gould, E. D., Weinberg, B. A., and Mustard, D. B. (2002). Crime rates and local labor market opportunities in the United States: 1979–1997. *Review of Economics and Statistics*, 84(1):45–61.

Grossman, H. I. and Kim, M. (1995). Swords or plowshares? A theory of the security of claims to property. *Journal of Political Economy*, 103(6):1275–1288.

Hansen, B., Miller, K., and Weber, C. (2017). The grass is greener on the other side: How extensive is the interstate trafficking of recreational marijuana? Technical report, National Bureau of Economic Research.

Hirshleifer, J. (1995). Anarchy and its breakdown. Journal of Political Economy, 103(1):26–52. Holland, J. (2010). The pot book: A complete guide to cannabis. Simon and Schuster.

Huber, A., Newman, R., and LaFave, D. (2016). Cannabis control and crime: Medicinal use, depenalization and the war on drugs. *B.E. Journal of Economic Analysis and Policy*, 16(4).

James, A. and Smith, B. (2017). There will be blood: Crime rates in shale-rich us counties. *Journal of Environmental Economics and Management*, 84:125–152.

Kilmer, B., Caulkins, J. P., Bond, B. M., and Reuter, P. (2010). Reducing Drug Trafficking Revenues and Violence in Mexico: Would Legalizing Marijuana in California Help? RAND Corporation, Santa Monica, CA.

Kuziemko, I. and Levitt, S. D. (2004). An empirical analysis of imprisoning drug offenders. *Journal of Public Economics*, 88(9-10):2043–2066.

Levitt, S. D. and Venkatesh, S. A. (2000). An economic analysis of a drug-selling gang's finances. *The Quarterly Journal of Economics*, 115(3):755–789.

Liao, D., Berzofsky, M., Heller, D., Barrick, K., and DeMichele, M. (2015). Treatment of missing data in the fbi's national incident based reporting system: A case study in the bakken region. In Joint Statistical Meetings, August, pages 8–13.

Lochner, L. (2004). Education, work, and crime: A human capital approach. *International Economic Review*, 45(3):811–843.

Lott, J. R. and Mustard, D. B. (1997). Crime, deterrence, and right-to-carry concealed handguns. *The Journal of Legal Studies*, 26(1):1–68.

Lynch, J. P. and Jarvis, J. P. (2008). Missing data and imputation in the uniform crime reports and the effects on national estimates. *Journal of Contemporary Criminal Justice*, 24(1):69–85.

Maltz, M. D. and Targonski, J. (2002). A note on the use of county-level UCR data. *Journal of Quantitative Criminology*, 18(3):297–318.

Mark Anderson, D., Hansen, B., and Rees, D. I. (2013). Medical marijuana laws, traffic fatalities, and alcohol consumption. *The Journal of Law and Economics*, 56(2):333–369.

Markowitz, S. (2005). Alcohol, drugs and violent crime. *International Review of Law and Economics*, 25(1):20–44.

Miron, J. A. (1999). Violence and U.S. prohibitions of drugs and alcohol. *American Law and Economics Review*, 1(1):78–114.

Morris, R. G., TenEyck, M., Barnes, J. C., and Kovandzic, T. V. (2014). The effect of medical marijuana laws on crime: evidence from state panel data, 1990-2006. *PloS One*, 9(3):e92816.

Musto, D. F. (1972). The marihuana tax act of 1937. *Archives of General Psychiatry*, 26(2):101–108.

National Association of Criminal Justice Data (NACJD) (2016). Uniform Crime Reporting Program Data: Offenses Known and Clearances by Arrest, United States, 2016 (37061) Codebook.

Pearl, B. (2018). Ending the war on drugs: By the numbers. ProCon.org (2019). 33 Legal Medical Marijuana State and DC.

Raphael, S. and Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law and Economics*, 44(1):259–283.

Rasmussen, D. W. and Benson, B. L. (1994). The economic anatomy of a drug war: Criminal justice in the commons. Rowman & Littlefield.

Resignato, A. J. (2000). Violent crime: a function of drug use or drug enforcement? *Applied Economics*, 32(6):681–688.

Reyes, J. W. (2007). Environmental policy as social policy? The impact of childhood lead exposure on crime. *The BE Journal of Economic Analysis & Policy*, 7(1).

Seaman, S., Galati, J., Jackson, D., and Carlin, J. (2013). What is meant by missing at random? Statistical Science, pages 257–268.

Targonski, J. R. (2011). A Comparison of Imputation Methodologies in the Offenses-Known Uniform Crime Reports. Dissertation, University of Illinois Chicago.

United and Nations (2018). World drug report - booklet 3. No. E.18.XI.9.

U.S. Federal Bureau of Investigation (FBI) (2020). Table 1 Data Declarations.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16(3):219–242.

Wen, H., Hockenberry, J. M., and Cummings, J. R. (2015). The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances. *Journal of Health Economics*, 42:64–80.

Tables

Table 1: Medical marijuana jurisdictions.

	Jurisdiction	Effective date		Jurisdiction	Effective date
1	California	11/6/1996	18	Connecticut	10/1/2012
2	Alaska	3/4/1998	19	District of Columbia	7/27/2010
3	Oregon	12/3/1998	20	Illinois	1/1/2014
4	Washington	11/3/1998	21	New Hampshire	7/23/2013
5	Maine	12/22/1999	22	Maryland	6/1/2014
6	Colorado	6/1/2001	23	Minnesota	5/30/2014
7	Hawaii	12/28/2000	24	New York	7/5/2014
8	Nevada	10/1/2001	25	Louisiana	5/19/2016
9	Montana	11/2/2004	26	Arkansas	11/9/2016
10	Vermont	7/1/2004	27	Florida	1/3/2017
11	Rhode Island	1/3/2006	28	North Dakota	4/18/2017
12	New Mexico	7/1/2007	29	Ohio	9/8/2016
13	Massachusetts	1/1/2013	30	Pennsylvania	4/17/2016
14	Michigan	12/4/2008	31	West Virginia	7/1/2019
15	Arizona	4/14/2010	32	Missouri	12/6/2018
16	New Jersey	6/1/2010	33	Oklahoma	6/26/2018
17	Delaware	7/1/2011	34	Utah	12/1/2018

Notes: Information obtained from ProCon.org (2019).

Table 2: Summary of previous literature.

	Alford (2014)	**Morris et al. (2014)**	**Huber et al. (2016)**	**Gavrilova et al. (2017)**	**Chu and Townsend (2019)**
UCR data	✓	✓	✓	✓	✓
DD	✓	✓	✓	✓	✓
DDD				✓	
3rd Difference				B/NB states	
Time Period	1995-2012	1990-2006	1970-2012	1994-2012	1988-2013
Jurisdiction	State	State	State	County	City
Model	OLS	OLS	WLS	WLS	OLS
MML Violent	≤ 0 (murder)	≤ 0	< 0	< 0 (border states)	$= 0$
MML Property	≤ 0; > 0	≤ 0	< 0		$= 0$

Notes: Difference-in-differences methods are typically used to analyze the effects of medical marijuana laws on violent and property crime. Gavrilova et al. (2017) extend this to a third difference and investigate the effects of MMLs in Mexican-border and non-border (B/NB) states. Time periods and levels of aggregation vary across the literature, as well as the use of ordinary least-squares (OLS) versus weighted least-squares (WLS).

Table 3: Patterns of missing data.

Murder	Robbery	Aggravated Assault	Burglary	Larceny	Motor Vehicle Theft	Freq.	Percentage
✓	✓	✓	✓	✓	✓	111,209	2.4701
✓	✓	✓	✓	✓		2,781	0.0618
✓	✓	✓	✓		✓	148	0.0033
✓	✓	✓	✓			22	0.0005
✓	✓	✓		✓	✓	112	0.0025
✓	✓	✓		✓		62	0.0014
✓	✓	✓			✓	4	0.0001
✓	✓	✓				8	0.0002
✓	✓		✓	✓	✓	2,190	0.0486
✓	✓		✓	✓		579	0.0129
✓	✓		✓		✓	6	0.0001
✓	✓		✓			16	0.0004
✓	✓			✓	✓	54	0.0012
✓	✓			✓		63	0.0014
✓	✓				✓	2	0.0000
✓	✓					18	0.0004
✓		✓	✓	✓	✓	17,048	0.3787
✓		✓	✓	✓		4,818	0.1070
✓		✓	✓		✓	65	0.0014
✓		✓	✓			138	0.0031
✓		✓		✓	✓	359	0.0080
✓		✓		✓		585	0.0130
✓		✓			✓	27	0.0006
✓		✓				117	0.0026
✓			✓	✓	✓	4,018	0.0892
✓			✓	✓		3,217	0.0715
✓			✓		✓	72	0.0016
✓			✓			235	0.0052
✓				✓	✓	379	0.0084
✓				✓		1,059	0.0235
✓					✓	44	0.0010
✓						512	0.0114

Notes: This table reports the pattern of missing agency-level crime data from the Uniform Crime Report (UCR). Violent crime rates include murder, robbery, and aggravated (agg.) assault. Burglary, larceny, and motor vehicle (M.V.) theft are property crimes. There are 64 combinations of missing data. ✓denotes the presence of the given crime rate in the pattern.

Table 4: Patterns of missing data continued.

Murder	Robbery	Aggravated Assault	Burglary	Larceny	Motor Vehicle Theft	Freq.	Percentage
	✓	✓	✓	✓	✓	476,207	10.5771
	✓	✓	✓	✓		60,865	1.3519
	✓	✓	✓		✓	386	0.0086
	✓	✓	✓			746	0.0166
	✓	✓		✓	✓	4,609	0.1024
	✓	✓		✓		5,763	0.1280
	✓	✓			✓	121	0.0027
	✓	✓				573	0.0127
	✓		✓	✓	✓	69,038	1.5334
	✓		✓	✓		37,347	0.8295
	✓		✓		✓	385	0.0086
	✓		✓			1,441	0.0320
	✓			✓	✓	4,953	0.1100
	✓			✓		10,627	0.2360
	✓				✓	295	0.0066
	✓					2,162	0.0480
		✓	✓	✓	✓	424,865	9.4368
		✓	✓	✓		276,376	6.1386
		✓	✓		✓	3,765	0.0836
		✓	✓			15,146	0.3364
		✓		✓	✓	32,149	0.7141
		✓		✓		90,994	2.0211
		✓			✓	2,744	0.0609
		✓				25,873	0.5747
			✓	✓	✓	258,327	5.7378
			✓	✓		427,357	9.4921
			✓		✓	9,455	0.2100
			✓			70,020	1.5552
				✓	✓	67,449	1.498124
				✓		345,745	7.679413
					✓	13,245	0.294187
						1,613,237	35.83194

Table 5: Crime rate sample sizes at each stage of the imputation and aggregation procedure.

Aggregation	agency-month	agency-year	county
1994	175,212	14,601	3,122
1995	176,052	14,671	3,124
1996	176,496	14,708	3,124
1997	177,396	14,783	3,124
1998	178,164	14,847	3,125
1999	179,892	14,991	3,125
2000	181,248	15,104	3,126
2001	181,908	15,159	3,126
2002	185,352	15,446	3,127
2003	186,624	15,552	3,127
2004	187,812	15,651	3,127
2005	188,604	15,717	3,127
2006	189,696	15,808	3,127
2007	191,136	15,928	3,127
2008	192,072	16,006	3,128
2009	193,068	16,089	3,128
2010	194,076	16,173	3,128
2011	194,688	16,224	3,128
2012	195,564	16,297	3,128
2013	195,504	16,292	3,128
2014	195,432	16,286	3,128
2015	195,324	16,277	3,128
2016	195,312	16,276	3,128
2017	195,312	16,276	3,128
Total	4,501,944	375,162	75,038

Table 6: Summary statistics for model variables.

variable	Source	N	Mean	STD	Min	Max
Pct. Male	Census	75,204	0.50	0.02	0.43	1.00
Pct. Age 10-19	Census	75,200	0.14	0.02	0.00	0.33
Pct. Age 20-24	Census	75,204	0.06	0.03	0.00	0.34
Pct. Black	Census	75,204	0.09	0.15	0.00	0.87
Pct. Hispanic	Census	75,204	0.06	0.11	0.00	0.98
Unemployment Rate	BLS	75,166	6.07	2.79	0.70	38.10
Population	Census	75,204	94,516.83	305,219.00	55.00	10,163,507.00
Violent Crime Rate	MI UCR	75,038	468.47	351.35	0.00	5,615.34
Property Crime Rate	MI UCR	75,038	4,183.22	3,583.40	0.00	73,228.36
Violent Crime Rate (GKZ Rep.)	MI UCR	58,834	499.49	369.48	0.00	5,615.34
Violent Crime Rate (GKZ)	NACJD	58,999	450.99	253.57	0.00	8,003.68
Violent Crime Rate	MI Simulated	75,038	489.00	350.16	0.00	5,832.21
Property Crime Rate	MI Simulated	75,038	4,349.84	3,684.34	0.00	71,354.25
Violent Crime Rate	MI Inflated 25%	75,038	464.86	340.06	0.00	6,168.57
Property Crime Rate	MI Inflated 25%	75,038	4,233.14	3,802.99	0.00	79,958.45
Violent Crime Rate	MI Inflated 50%	75,038	463.12	333.59	0.00	5,802.97
Property Crime Rate	MI Inflated 50%	75,038	4,294.74	3,968.28	0.00	74,924.83

Notes: Crime rate means are population weighted. This table reports county-year data. Demographics and population come from the United States (U.S.) Census Bureau and the U.S. Bureau of Labor Statistics (BLS). Crime rate data come from the Uniform Crime Report (UCR) and were imputed and aggregated from agency-level to county-level following the procedures described in this paper.

Table 7: Estimation results using multiply imputed data.

Violent	(1)	(2)	(3)	(4)	(5)
MML	-40.629***	-34.032***	-34.281***	-28.779***	-29.102***
	(5.777)	(6.071)	(6.367)	(6.529)	(6.600)
Border State MML		-71.397***		-65.437***	-61.672**
		(20.070)		(20.316)	(24.473)
Urban MML			-36.573***	-33.442***	-31.347**
			(11.945)	(12.035)	(13.325)
Urban Border MML					-12.3
					(30.849)
Constant	1086.742***	1081.611***	1054.073***	1052.167***	1051.202***
	(212.250)	(212.209)	(212.155)	(212.140)	(212.072)
R-squared	0.701	0.706	0.706	0.706	0.706
Number of Counties	3121	3121	3121	3121	3121
Property	(1)	(2)	(3)	(4)	(5)
MML	-240.187***	-211.801***	-201.951***	-179.224**	-180.24**
	(71.129)	(78.494)	(72.977)	(79.004)	(79.612)
Border State MML		-307.233**		-270.276**	-258.462*
		(133.389)		(136.941)	(149.116)
Urban MML			-220.302*	-207.373	-200.799
			(132.144)	(133.881)	(156.373)
Urban Border MML					-38.593
					(228.046)
Constant	22710.334***	22688.256***	22513.545***	22505.672***	22502.646***
	(2207.453)	(2207.176)	(2202.738)	(2202.750)	(2203.547)
R-squared	0.582	0.587	0.587	0.587	0.587
Number of Counties	3121	3121	3121	3121	3121

Notes: Times series spans 1994 to 2017. *,**,*** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 8: Estimation results for comparison of results using NACJD data from Gavrilova et al. (2017) and our imputation procedure.

Violent (MI)	(1)	(2)	(3)	(4)	(5)
MML	-48.705***	-34.352***	-40.62***	-28.526**	-28.929**
	(9.637)	(11.065)	(10.627)	(11.602)	(11.836)
Border State MML		-64.477***		-58.253**	-55.773**
		(22.428)		(22.835)	(27.513)
Urban MML			-41.777**	-37.263**	-34.691*
			(16.909)	(17.082)	(20.442)
Urban Border MML					-8.973
					(37.349)
Constant	701.98***	693.991***	675.78***	671.393***	671.267***
	(242.564)	(242.500)	(242.614)	(242.528)	(242.524)
R-squared	0.718	0.724	0.724	0.724	0.724
Number of Counties	3098	3098	3098	3098	3098
Violent (GKZ 2017)	(1)	(2)	(3)	(4)	(5)
MML	-28.035*	-11.142*	-23.695*	-8.536*	-7.457*
	(7.933)	(6.161)	(8.945)	(6.625)	(6.705)
Border State MML		-76.054***		-73.259**	-79.973**
		(28.106)		(29.139)	(35.711)
Urban MML			-22.498	-16.725	-23.637
			(14.140)	(14.918)	(16.105)
Urban Border MML					24.281
					(37.277)
Constant	225.172***	219.553*	222.078***	217.459	217.112
	(7.933)	(6.161)	(8.945)	(6.625)	(6.705)
R-squared	0.74	0.742	0.741	0.742	0.742
Number of Counties	3100	3100	3100	3100	3100

Notes: Imputations are conducted using population data from the Uniform Crime Report to calculate crime rates in the imputation step. *,**,*** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 9: Estimation results using simulated multiply imputed data.

Violent	(1)	(2)	(3)	(4)	(5)
MML	-46.336***	-40.284***	-40.788***	-35.715***	-36.293***
	(5.720)	(6.055)	(6.166)	(6.386)	(6.426)
Border State MML		-65.506***		-60.323***	-53.609**
		(18.273)		(18.487)	(22.172)
Urban MML			-31.968***	-29.082***	-25.346**
			(10.534)	(10.610)	(11.635)
Urban Border MML					-21.933
					(28.039)
Constant	1264.979***	1260.272***	1236.423***	1234.666***	1232.946***
	(178.977)	(178.921)	(178.785)	(178.757)	(178.707)
R-squared	0.721	0.727	0.727	0.727	0.727
Number of Counties	3121	3121	3121	3121	3121
Property	(1)	(2)	(3)	(4)	(5)
MML	-459.477***	-457.5***	-468.342***	-465.755***	-475.239***
	(86.727)	(95.839)	(90.128)	(97.505)	(98.532)
Border State MML		-21.403		-30.768	79.499
		(149.800)		(154.762)	(164.989)
Urban MML			51.077	52.549	113.903
			(152.420)	(154.134)	(182.089)
Urban Border MML					-360.2
					(233.837)
Constant	25130.876***	25129.338***	25176.502***	25175.605***	25147.363***
	(2345.349)	(2344.584)	(2349.727)	(2349.380)	(2349.436)
R-squared	0.566	0.571	0.571	0.571	0.571
Number of Counties	3121	3121	3121	3121	3121

Notes: These data are imputed based on a simulated data set to test the MAR assumption. Times series spans 1994 to 2017. *,**,*** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 10: Estimation results using multiply imputed data. Imputed values are inflated by 25%.

Violent	(1)	(2)	(3)	(4)	(5)
MML	-34.684***	-27.515***	-27.178***	-21.25***	-21.397***
	(5.764)	(6.004)	(6.297)	(6.395)	(6.461)
Border State MML		-77.594***		-70.488***	-68.788***
		(21.364)		(21.692)	(26.360)
Urban MML			-43.248***	-39.876***	-38.93***
			(12.384)	(12.516)	(13.772)
Urban Border MML					-5.554
					(33.080)
Constant	1148.617***	1143.041***	1109.985***	1107.932***	1107.496***
	(218.753)	(218.738)	(218.714)	(218.723)	(218.616)
R-squared	0.665	0.669	0.669	0.669	0.669
Number of Counties	3121	3121	3121	3121	3121
Property	(1)	(2)	(3)	(4)	(5)
MML	-179.263**	-145.851*	-146.887*	-119.036	-121.277
	(77.518)	(85.668)	(79.437)	(86.133)	(86.750)
Border State MML		-361.633**		-331.213**	-305.159*
		(147.002)		(151.115)	(165.411)
Urban MML			-186.538	-170.694	-156.197
			(145.109)	(146.948)	(170.704)
Urban Border MML					-85.109
					(251.242)
Constant	24723.461***	24697.473***	24556.832***	24547.183***	24540.51***
	(2709.563)	(2709.212)	(2712.755)	(2712.648)	(2713.443)
R-squared	0.481	0.486	0.486	0.486	0.486
Number of Counties	3121	3121	3121	3121	3121

Notes: Imputed values are inflated by 25% to test robustness of the MAR assumption. Times series spans 1994 to 2017. *,**,*** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 11: Estimation results using multiply imputed data. Imputed values are inflated by 50%.

Violent	(1)	(2)	(3)	(4)	(5)
MML	-30.11***	-22.767***	-21.757***	-15.744**	-15.802**
	(6.019)	(6.211)	(6.512)	(6.551)	(6.612)
Border State MML		-79.473***		-71.505***	-70.832**
		(22.530)		(22.949)	(28.003)
Urban MML			-48.13***	-44.71***	-44.335***
			(12.826)	(13.008)	(14.217)
Urban Border MML					-2.197
					(35.351)
Constant	1230.35***	1224.639***	1187.357***	1185.274***	1185.102***
	(241.525)	(241.553)	(241.489)	(241.537)	(241.394)
R-squared	0.633	0.638	0.638	0.638	0.638
Number of Counties	3121	3121	3121	3121	3121
Property	(1)	(2)	(3)	(4)	(5)
MML	-107.401	-66.864	-67.072	-33.372	-34.704
	(83.379)	(92.028)	(86.059)	(93.178)	(93.826)
Border State MML		-438.762***		-400.768**	-385.291**
		(159.589)		(163.003)	(180.739)
Urban MML			-232.365	-213.194	-204.582
			(151.450)	(153.154)	(176.803)
Urban Border MML					-50.558
					(274.077)
Constant	26720.879***	26689.349***	26513.314***	26501.64***	26497.675***
	(3308.020)	(3308.359)	(3318.729)	(3318.845)	(3319.560)
R-squared	0.43	0.436	0.436	0.436	0.436
Number of Counties	3121	3121	3121	3121	3121

Notes: Imputed values are inflated by 50% to test robustness of the MAR assumption. Times series spans 1994 to 2017. *,**,*** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figures

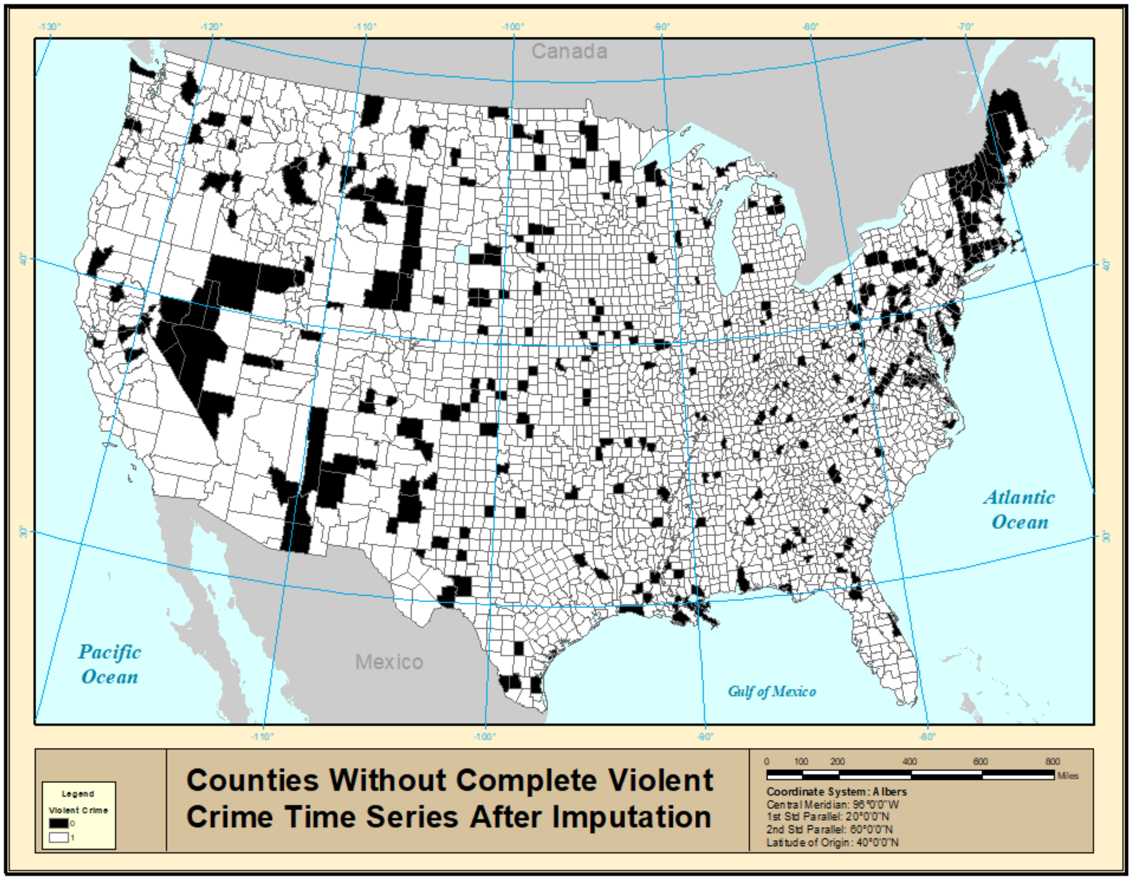

Figure 1: Geographic distribution of missing crime data for crime estimation using county aggregated UCR data that has been multiply imputed.