**Effect of Marijuana legalization on Violent Crimes**

# **Abstract**

This project meticulously explores the intricate relationship between marijuana legalization and violent crime rates in the United States, utilizing advanced econometric techniques and a comprehensive set of demographics, political, and religious variables. Analysing data from 2014 and 2019 and employing propensity score matching, logistic regression, and inverse probability weighted regression adjustment (IPWRA), our research sheds light on the societal repercussions of marijuana policy changes.

The key findings indicate that states embracing marijuana law liberalization, on average, experienced a notable reduction in violent crime rates compared to non-liberalized states. Our treatment model, guided by logistic regression and rigorous variable selection, identifies influential factors such as political ideology, religious affiliation, and unemployment rates that shape states' decisions regarding marijuana law reforms. The outcome model, incorporating crucial covariates, provides a nuanced understanding of the complex relationship between marijuana policy shifts and violent crime rates in 2019.

Notably, the estimation of treatment effects through simpler propensity scores and IPWRA consistently underscores the reduced effect of marijuana law liberalization on violent crime rates. Average Treatment Effect on the Untreated, Average Treatment Effect on the Treated, and Average Treatment Effect collectively support the conclusion that adopting marijuana law liberalization is associated with a statistically significant reduction in violent crime rates.

These empirically grounded insights significantly contribute to the ongoing discourse surrounding marijuana policy changes and offer a nuanced perspective on their potential societal consequences. Policymakers, researchers, and the public can use these findings to make informed decisions and engage in discussions regarding the impact of drug policy shifts on public safety.

# **Introduction:**

The evolving landscape of marijuana legislation in the United States has generated substantial debate regarding its societal implications, particularly in relation to violent crime rates. As attitudes toward cannabis shift and legalization becomes more widespread across states, understanding the nuanced impact on public safety becomes increasingly crucial. This project aims to address specific gaps in the existing literature and contribute to a deeper understanding of the intricate dynamics between marijuana law liberalization and violent crime, employing advanced econometric techniques and a rich array of demographic, political, and religious variables.

While previous research has explored the broad consequences of marijuana liberalization, there exists a notable gap in understanding the nuanced relationships between evolving drug policies and violent crime rates at the state level. Our analysis seeks to fill this gap by employing a comprehensive treatment selection model that incorporates variables such as gender distribution, age demographics, religious affiliation, and unemployment rates. By doing so, I aim to address the limitations of prior research and provide a more detailed examination of the potential impacts of marijuana law changes on public safety.

To guide this investigation, the following research questions are explicitly stated:

1. How does marijuana law liberalization influence violent crime rates at the state level?
2. What are the specific demographic, political, and religious factors that contribute to states adopting marijuana law liberalization?
3. To what extent do these factors interact, and how do they shape the overall relationship between marijuana policy changes and violent crime rates?

By formulating these research questions, this study seeks to provide a clear roadmap for the analysis and contribute empirically grounded insights to the ongoing discourse on the societal consequences of marijuana policy changes. Through advanced statistical methods and a thoughtful selection of covariates, the goal is to fill existing gaps in the literature and offer a more comprehensive understanding of the multifaceted dynamics involved in the relationship between marijuana legalization and violent crime rates in the United States.

# **Literature review**

The intricate relationship between marijuana liberalization and violent crime rates has been extensively explored within the existing body of literature. While numerous studies contribute valuable insights, a critical analysis reveals certain contradictions and limitations that necessitate a nuanced understanding of this complex interplay.

Felson, Adamczyk, and Thomas (2019)[1](#one) emphasize the influential role of media framing and declining religious affiliation in shaping public opinion on marijuana. Their insights provide a foundation for understanding societal attitudes, yet the limitations lie in the generalizability of these findings, as media effects can vary across demographic and regional contexts. Miech and Koester's (2012)[2](#Two) analysis challenges conventional perceptions of marijuana use trends by underscoring the significance of period effects across cohorts. However, a critical examination raises questions about the potential influence of unobserved variables and cohort-specific factors that may impact the generalizability of their conclusions.

In the exploration of state-level cannabis legalization, Spetz et al. (2019)[3](#Three) provide a comprehensive framework by delving into the social and political factors that drive such policy changes. While their study contributes valuable insights, it is essential to critically evaluate the transferability of their findings to diverse state contexts, considering the variability in political landscapes and public attitudes. Dragone, Prarolo, Vanin, and Zanella (2019)[4](#Four) specifically focus on the impact of recreational marijuana legalization on crime rates, revealing a reduction in rapes and property crimes in legalized regions. The findings highlight a potential positive outcome, yet the study's limitations may include the exclusion of other crime categories and the need for a more granular analysis of specific legalization models. Insights from the study by Burdette et al. (2018)[5](#Five) revealed intriguing patterns in marijuana use for medical and recreational purposes, highlighting the role of religious involvement and health status. While valuable, the study's limitations in addressing the diverse religious landscape and potential confounders underscore the need for a more comprehensive analysis.

These studies collectively underscore the need for a nuanced analysis, considering socio-economic, cultural, and political dimensions in understanding the interplay between marijuana liberalization and crime rates. Our research builds upon this rich foundation, aiming to contribute empirical insights through a rigorous examination of state-level data. Employing advanced statistical methods, I seek to disentangle the complex dynamics involved, ensuring our research is situated within the broader landscape of literature and provides a comprehensive understanding of the factors influencing the relationship between marijuana policy and violent crime rates.

# **Data**

The dataset assembled for this study incorporates a thoughtfully curated selection of variables aimed at capturing the complex dynamics surrounding the adoption of marijuana policy and its potential repercussions on crime rates. An instrumental component of our Treatment Selection Model is the Historical Marijuana Policy Stance, representing the historical position each state held regarding marijuana policies and serving as a crucial determinant in this investigation.

Political Ideology, a pivotal factor influencing the likelihood of adopting marijuana liberalization, is sourced from the Pew Research Centre’s comprehensive study on political ideology distribution[6](#Six) within each state. In tandem, Religious Affiliation data from the same source[7](#Seven) provides insight into the percentage of the state population affiliated with major religious groups, contributing to the mosaic of attitudes towards marijuana liberalization. Complementing these factors, the Unemployment Rate, obtained from the Bureau of Labor Statistics[8](#Eight), adds an economic dimension to our Treatment Selection Model. This variable reflects the percentage of the labour force unemployed in each state, potentially influencing the policy landscape.

Transitioning to the Outcome Model Control Variables, I delve into demographic intricacies. Demographic Data, focusing on Age Distribution and Gender Distribution, is derived from the U.S. Census Bureau[9](#Nine). These variables, essential for understanding population dynamics, provide context to potential impacts on crime rates. Education Attainment, a critical societal metric, is represented as the Percentage of Education Attainment and is sourced from the National Centre for Education Statistics[10](#Ten). Additionally, the Average State Poverty Rate, drawn from the U.S. Census Bureau[12](#twelve), captures the economic landscape, contributing to a holistic understanding of potential influencing factors.

Given the challenges in accessing data for the year 1995, the approach pivoted towards a robust examination of 2014 and 2019. This strategic shift allows for a comprehensive exploration of the evolving landscape surrounding marijuana policy and its potential ramifications on crime rates. Through meticulous data collection from reputable sources, a rigorous and well-informed analysis is ensured, providing a solid foundation for this study.

# **Analysis and Results**

## **Treatment Selection Model**

### **Propensity Score Method**

The approach to understanding the impact of marijuana law liberalization involves the use of propensity scores. These scores, representing the probability of a state adopting liberalization based on observed covariates, play a pivotal role in minimizing selection bias. By estimating the likelihood of treatment assignment, I can ensure comparability between treated and control groups.

### **Implementation of Propensity Score with Logistic Regression:**

Logistic regression is a robust statistical method which is employed to derive the propensity scores. This regression model incorporated key covariates such as historical marijuana policy stance, political ideology, religious affiliation, and the unemployment rate. Logistic regression, being well-suited for binary outcomes, provided insights into the probability of adopting marijuana liberalization.

The logistic regression equation is expressed as:

Here, P (Y = 1) represents the probability of adopting marijuana legislation.

### **Treatment selection Model Sketch and DAG:**

The variable selection logic focused on key determinants of marijuana law liberalization and potential confounders. Historical marijuana policy stance, political ideology, religious affiliation, and the unemployment rate were chosen for their perceived influence on policy decisions. The Directed Acyclic Graph visually represents the causal relationships, illustrating how these variables interplay.

Age distribution

Gender distribution

Unobserved

Marijuana Legalization

Unemployment rate

Religious affiliation

Political ideology

DAG 1. DAG representing Treatment model.

The DAG serves as a graphical depiction of the causal pathways and dependencies between Gender distribution (Male and Female percentages in 2014 and 2019), age distribution (Young, Middle-aged, and Old percentages in 2014 and 2019), religious affiliation (Catholic, Evangelical Protestant, Historically Black Protestant, Jewish, Mainline Protestant, and Unaffiliated percentages), unemployment rates in 2014 and 2019, and political ideology (Conservative, Liberal) and the propensity for states to adopt marijuana legalization. Arrows in the DAG illustrate the assumed causal relationships, elucidating how changes in one variable may influence others, ultimately impacting the likelihood of marijuana policy liberalization. Unobserved factors, crucial yet challenging to measure directly, were also incorporated into the DAG, acknowledging their potential influence on the observed variables and the ultimate adoption of marijuana legalization. This graphical representation not only guides our model construction but also facilitates a deeper understanding of the intricate web of factors contributing to the state-level decisions surrounding marijuana policy changes.

### **Logistic regression models**

Four logistic regression models (Model 1 to Model 4) were constructed to discern the factors influencing states' marijuana policy stances. These models incorporated variables such as gender distribution, age demographics, religious affiliation, and political ideology, providing insights into the determinants of treatment assignment. The selection process considered several key factors, including Area Under the Curve (AUC) values, type 1 and type 2 errors, and the overall interpretability of coefficients.

Table 1: Logistic regression model results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model1 | Model2 | Model3 | Model4 |
| Male\_2014 | 5196.888 | 120.577 | 145.676 | -116.712 |
|  | (0.000) | (277.580) | (127.493) | (119.715) |
|  |  | 0.66 | 0.25 | 0.33 |
| Young\_2014p | -1.06e+04 | -147.441 | -62.201 | -12.322 |
|  | (0.000) | (181.336) | (81.605) | (88.154) |
|  |  | 0.42 | 0.45 | 0.89 |
| Middle\_aged\_2014p | -2.12e+04 |  |  |  |
|  | (0.000) |  |  |  |
| unemployment\_2014\_s | -110.823 | -3.035 | -1.408 | -9.155 |
|  | (0.000) | (5.889) | (3.286) | (4.319) |
|  |  | 0.61 | 0.67 | 0.03 |
| Catholic | 866.216 | -27.012 | -29.554 | -5.178 |
|  | (0.000) | (26.375) | (12.640) | (13.917) |
|  |  | 0.31 | 0.02 | 0.71 |
| Liberal | 21252.758 |  |  |  |
|  | (0.000) |  |  |  |
| Evangelical Protestant | 746.405 | 38.032 |  | 36.662 |
|  | (0.000) | (22.623) |  | (13.975) |
|  |  | 0.09 |  | 0.01 |
| Historically Black Protestant | 783.288 | -5.018 |  |  |
|  | (0.000) | (40.429) |  |  |
|  |  | 0.90 |  |  |
| Mainline Protestant | 270.107 | -8.079 |  |  |
|  | (0.000) | (22.671) |  |  |
|  |  | 0.72 |  |  |
| Unaffiliated | 2154.084 | -73.152 | -57.876 |  |
|  | (0.000) | (54.064) | (23.370) |  |
|  |  | 0.18 | 0.01 |  |
| relig\_att\_week | 1999.557 |  |  |  |
|  | (0.000) |  |  |  |
| relig\_att\_year | 1295.605 | 3.630 |  |  |
|  | (0.000) | (27.953) |  |  |
|  |  | 0.90 |  |  |
| Intercept | -1831.219 | -8.993 | -37.217 | 55.938 |
|  | (0.000) | (146.271) | (47.377) | (59.312) |
|  |  | 0.95 | 0.43 | 0.35 |
| Pseudo R-squared | 1.00 | 0.77 | 0.67 | 0.69 |

Note: In each cell, the first number is the odds ratio, the second in parentheses its

standard error, and the third is the associated p-value for a test of the null hypothesis of no effect.

Table : Confusion Matrices for Models 1-4

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Treat | Prediction | | | | | | | |
| Model1 | | Model2 | | Model3 | | Model4 | |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 0 | 9 | 15 | 24 | 0 | 23 | 1 | 22 | 2 |
| 1 | 0 | 19 | 2 | 17 | 2 | 17 | 2 | 17 |

Model 3, the selected treatment model, revealed significant insights into the factors guiding states' decisions on marijuana liberalization. Conservative states exhibited a lower likelihood of adopting marijuana liberalization, as evidenced by the negative coefficient for Conservative. The influence of religious factors was evident, with increased Catholic affiliation associated with a reduced likelihood of marijuana liberalization (Catholic coefficient). The impact of religious affiliation on treatment assignment was nuanced, as highlighted by the positive coefficients for Evangelical Protestant and Unaffiliated. Unemployment\_2014\_s demonstrated a notable effect, suggesting that states with higher unemployment rates in 2014 were more likely to adopt marijuana liberalization. Religious attendance (Relig\_Att\_Week) and year-long religious salience (Relig\_Att\_Year) played contrasting roles, with the former reducing the likelihood and the latter increasing the likelihood of marijuana liberalization.

Model 3's selection as the treatment model was informed by a comprehensive evaluation of alternative specifications. Model 1, characterized by a pseudo R-squared value of 1.00, raised concerns about overfitting and lack of generalizability. Model 2 faced challenges with variable significance, and Model 4 exhibited similar issues alongside a marginally lower pseudo R-squared value.

Propensity scores derived from Model 3 facilitated a closer examination of treatment assignment probabilities. The Receiver Operating Characteristic (ROC) curve demonstrated the model's ability to distinguish between treatment and control groups, with an AUC value confirming satisfactory predictive accuracy.

### **Confusion Matrices and Type 1/Type 2 Errors:**

Confusion matrices were instrumental in assessing the model's classification performance. Model 3's balanced trade-off between type 1 and type 2 errors contributed to its selection as the treatment model. The decision was based on minimizing false positives (assigning treatment incorrectly) and false negatives (failing to assign treatment when appropriate).

### **Propensity Score Distribution: Treatment vs. Control**

A graph of a patient

Description automatically generatedTo evaluate the balance achieved after calculating propensity scores, a histogram was constructed to illustrate the distribution of propensity scores for both the treatment and control groups. The histogram portrays the frequency distribution of propensity scores, offering a visual representation of the overlap between the two groups. In the histogram, the treatment group is depicted in a shaded colour, while the control group is outlined in black. Although the overlap is not extensive, some degree of common support is evident, indicating that there are units in both groups with similar and overlapping propensity scores. The histogram serves as a preliminary visual assessment of the balance achieved through propensity score matching, providing a foundation for further analyses and ensuring the suitability of the matched samples for subsequent treatment effect estimation.

Fig 1. Overlap in propensity scores for treatment and control groups

### **Covariate Overlap Scatter Plots: Examining Two-Dimensional Overlap**

To further assess the covariate overlap between the treatment and control groups, scatter plots were generated for key covariates.

A graph of a number of objects

Description automatically generated with medium confidence**Covariate Overlap in Unemployment Rate and Religious Affiliation:**

The scatter plot on the right-hand side illustrates the overlap in unemployment rates and the percentage of Catholics for both the treatment and control groups. The 'Treatment' group is denoted by triangles (△), while the 'Control' group is represented by circles (○). The plot illustrates a considerable overlap between the two groups, indicating that, despite variations, there are states in both categories with similar unemployment rates and percentages of Catholics. This overlap is crucial for the effectiveness of the inverse probability weighted regression adjustment, providing a foundation for unbiased treatment effect estimation.

Fig 2. Overlap between covariates

A graph of a number of people

Description automatically generated with medium confidence

Fig 3. Overlap between Covariates

**Covariate Overlap in Religious Affiliation and Young Population:**

In a similar vein, the scatter plot explores the overlap in the percentage of Catholics and the percentage of young individuals in the population between the treatment and control groups, with triangles (△) indicating the 'Treatment' group and squares (□) representing the 'Control' group. The plot demonstrates a notable overlap, indicating that states with marijuana law liberalization and those without share common ground concerning the religious affiliation and age distribution of their populations. This covariate overlap reinforces the validity of the inverse probability weighted regression adjustment, as it suggests that observed units from both groups exhibit similar characteristics in terms of religious composition and age distribution.

## **Outcome Model in Regression Adjustment Context:**

The outcome model explores the intricate relationship between marijuana law liberalization and violent crime rates, focusing on the year 2019 as the reference point.

### **Outcome model sketch and DAG:**

The DAG for the outcome model illustrates the assumed causal relationships among covariates and the outcome variable (violent crime rates in 2019). The baseline violent crime rates in 2014 act as a key confounder, connecting to both the treatment variable and the outcome. This covariate is crucial for controlling the pre-existing differences in crime rates, ensuring a more accurate estimation of the causal impact of marijuana law liberalization on violent crime rates in 2019.Covariates include demographic factors, such as the percentage of the percentage of the male population in 2019 is connected to the outcome variable, reflecting the potential influence of gender demographics on violent crime rates. Similarly, the percentage of the young population in 2019, median income in 2019, religious attendance, percentage of poverty in 2019 are linked to the outcome, emphasizing the importance of each covariate in understanding crime dynamics. The unemployment rate in 2019 serves as another crucial covariate, capturing economic factors that may contribute to variations in crime rates.

Religious attendance (relig\_att\_week, relig\_att\_year)

Percentage of Poverty in 2019 (poverty\_2019)

Median Income (Income\_2019)

Violent Crimes in 2014 (crime\_2014)

Violent Crimes in 2019 (crime\_2019)

Unemployment rates (unemployment\_2019)

Gender distribution (Male\_2019)

Unobserved

Age distribution (Young\_2019)

DAG 2: DAG representing Outcome Model.

The arrows in the DAG represent the assumed causal pathways, guiding the estimation of the regression model. The outcome model's construction adheres to sound causal inference principles, ensuring that the chosen covariates effectively control for potential confounding factors, ultimately contributing to a robust analysis of the relationship between marijuana law liberalization and violent crime rates in 2019.

The DAG for the outcome model considers unobserved variables to account for potential omitted factors that might affect both marijuana law liberalization and violent crime rates. While these unobserved variables are not directly measured, their influence is acknowledged to ensure a more comprehensive understanding of the causal relationships.

The regression model is estimated with the specified covariates, providing insights into the impact of marijuana law liberalization on violent crime rates in 2019. The discussion interprets the coefficients, their statistical significance, and the overall explanatory power of the model. By examining the relationship between the treatment variable and the outcome, this analysis aims to discern the potential causal effect of marijuana law liberalization on violent crime rates in the selected year, considering the nuanced influence of unemployment rates in 2019.

### **Estimation of Treatment Effects:**

The estimation of treatment effects encompasses two approaches: the Inverse Probability Weighted Regression Adjusted method and simpler propensity score matching. The IPWRA approach incorporates weights derived from estimated propensity scores, providing a nuanced understanding of the treatment effect while addressing varying probabilities of treatment assignment. This method mitigates potential bias and ensures a robust estimation of the causal impact.

Additionally, propensity score matching is employed to create balanced pairs of treated and control units with similar propensity scores. This approach facilitates a direct comparison of outcomes, specifically focusing on the average treatment effect (ATE).

### **Estimating Treatment Effects using simpler propensity scores:**

In the analysis of the outcome model with regression adjustment using inverse probability weighting (IPW), separate regressions were conducted for the control and treatment groups. This study delved into Average Treatment Effect on the Untreated (ATU) and Average Treatment Effect on the Treated (ATT), leveraging propensity scores to discern the nuanced impact of policy changes.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Summary of teffect | | |
| treat | Mean | Std. dev. | Freq. |
| 0 | -.1302025 | 0.0812758 | 24 |
| 1 | -.0838801 | 0.0878563 | 19 |
| Total | -.1097345 | 0.0864133 | 43 |

Table : Summary of treatment effect for regression adjustment model using IPW

ATU measures the average impact of marijuana law liberalization on non-liberalized states, offering insights into whether the policy change influences regions that have not undergone liberalization. The calculated ATU revealed a mean treatment effect of -0.1302 with a standard deviation of 0.0813. This indicates that, on average, non-liberalized states experienced a reduction in violent crime rates following marijuana law liberalization. ATT focuses on the average impact of marijuana law liberalization on treated or liberalized states. It gauges whether the policy change has a discernible effect on regions that have embraced marijuana law reforms. The analysis yielded an ATT mean treatment effect of -0.0839 with a standard deviation of 0.0879. This suggests that, on average, liberalized states witnessed a reduction in violent crime rates compared to what would be expected in the absence of marijuana law changes. The overall ATE represents the difference in propensity scores between the treated and control groups. In our study, the ATE demonstrated a mean treatment effect of -0.1097 with a standard deviation of 0.0864. This comprehensive measure encapsulates the average impact of marijuana law liberalization on violent crime rates across all states, providing a holistic perspective on the policy's potential influence.

These findings collectively suggest a consistent trend across ATU, ATT, and ATE, indicating a mitigating effect of marijuana law liberalization on violent crime rates. The negative mean treatment effects across these metrics imply that regions with liberalized marijuana policies experienced, on average, a reduction in violent crime rates compared to non-liberalized states. This insight contributes to a nuanced understanding of the intricate relationship between drug policy changes and crime dynamics, shedding light on potential societal benefits associated with marijuana law reforms.

### **Estimating Treatment Effects using Inverse Probability Weighted Regression Adjustment:**

The exploration of treatment effects through simpler propensity scores and IPWRA provides valuable insights into the potential impact of marijuana law liberalization on violent crime rates. In the simpler propensity score matching, I observed nuanced relationships between crime rates and various covariates. Notably, poverty rates in 2019 exhibited a positive association in non-liberalized states, albeit marginally significant. In liberalized states, the negative coefficient for poverty rates, although not statistically significant, hinted at a potential mitigating effect.

When employing IPWRA, the treatment effect estimation revealed a statistically significant negative coefficient for the treatment variable (treat), reinforcing the notion that states adopting marijuana law liberalization experienced, on average, a reduction in violent crime rates. The 95% confidence interval further substantiates the robustness of this finding.

The ATE serves as a crucial metric in understanding the overall impact of marijuana law liberalization on violent crime rates. In our analysis, ATE signifies the average difference in crime rates between liberalized and non-liberalized states, considering all relevant covariates. The negative ATE value(-.1105853) supports the conclusion that, on average, adopting marijuana law liberalization is associated with a reduction in violent crime rates. The POmean (.3211547) analysis provided additional depth, indicating a lower average violent crime rate in states with marijuana law liberalization compared to non-liberalized states. This suggests a potential preventive effect of marijuana policy changes on violent crime rates.

The utilization of both simpler propensity scores and IPWRA enhances the reliability of our findings. Propensity score matching mitigates selection bias, ensuring a more balanced comparison between treated and control groups. Meanwhile, IPWRA considers the intricacies of treatment assignment probabilities, offering a comprehensive understanding of the causal relationship between marijuana law liberalization and violent crime rates.

# **Conclusion**

In conclusion, this comprehensive analysis has systematically addressed the research questions and objectives, providing a nuanced understanding of the intricate relationship between marijuana law liberalization and violent crime rates in the United States. Leveraging a sophisticated methodology that combines propensity score matching, logistic regression, and IPWRA, and have endeavoured to unravel the nuanced dynamics that underlie the evolving landscape of drug policy and its potential impact on societal outcomes. The main findings of our analysis consistently indicate a mitigating effect of marijuana law liberalization on violent crime rates.

## **Key Findings:**

**Reduction in Violent Crime Rates**: The analysis consistently reveals that states embracing marijuana law liberalization experienced a notable reduction in violent crime rates compared to non-liberalized states. The consistency of this finding across different treatment effect estimation methods, including simpler propensity scores and IPWRA, strengthens its credibility and it directly aligns with the primary research question, providing clear insights into the influence of marijuana law liberalization on crime rates at the state level.

**Influential Factors**: The treatment model, guided by logistic regression and rigorous variable selection, identified influential factors shaping states' decisions on marijuana law reforms. Political ideology, religious affiliation, and unemployment rates emerged as crucial determinants, directly addressing the research objective of understanding specific demographic, political, and religious factors contributing to states adopting marijuana law liberalization.

**Empirical Grounding**: By addressing specific gaps in the existing literature and formulating explicit research questions, this analysis contributes to a deeper understanding of the complex relationships between marijuana policy changes and violent crime rates at the state level. While our findings indicate a potential preventive effect of marijuana policy shifts on violent crime rates, it is crucial to acknowledge the multifaceted nature of these dynamics. Societal, economic, and law enforcement factors may interact with marijuana policy changes, influencing their impact on crime rates in complex ways.

In essence, this project enhances understanding of the intricate dynamics between marijuana legalization and violent crime rates, providing a nuanced perspective on their potential societal consequences. The findings offer a valuable foundation for further exploration in the evolving landscape of drug policy, contributing to evidence-based decision-making by policymakers, researchers, and the public. As marijuana legislation continues to evolve, these insights provide a thoughtful and comprehensive examination of the potential consequences of such legal changes.

# **Appendix A: Do file for the analysis**

clear

log using "Econometrics\_final\_project", text replace

\*set working directory

cd "D:\Masters\Sem4\Econometrics\Project"

import delimited "D:\Masters\Sem4\Econometrics\Project\data\Final\_data.csv" // import the data

\*Data Exploration

\* Generate the "treat" variable

gen treat = .

\* Set "treat" to 0 for states with MML before 2016

replace treat = 0 if firstyearmml < 2016

\* Set "treat" to 1 for states with MML is missing

replace treat = 1 if missing(firstyearmml) & treat != 0

\* Drop states where "treat" is missing

drop if missing(treat)

\*Converting string variable into numerical variable

gen crime\_2014\_no\_commas = subinstr(violent\_crimepc\_2014, ",", "", .)

gen crime\_2014 = real(crime\_2014\_no\_commas)

gen crime\_2019\_no\_commas = subinstr(violent\_crimespc\_2019, ",", "", .)

gen crime\_2019 = real(crime\_2019\_no\_commas)

gen income\_2019\_no\_commas = subinstr(income\_2019, ",", "", .)

gen income\_2019\_n = real(income\_2019\_no\_commas)

drop income\_2019\_no\_commas income\_2019

\*Summary statistics for all the variables in treatment model

summarize male\_2014 young\_2014p middle\_aged\_2014p ///

unemployment\_2014 conservative catholic liberal catholic ///

evangelicalprotestant historicallyblackprotestant jewish ///

mainlineprotestant unaffiliated relig\_att\_week relig\_att\_year

\* Scaling variable (min-max)

gen unemployment\_2014\_s = (unemployment\_2014 - 3.5) / (9.2 - 3.5)

sum unemployment\_2014\_s

\*Summary statistics for all the variables in outcome model

summarize crime\_2019 povertyp\_2019 edu\_att\_2019 income\_2019\_n ///

male\_2019 unemployment\_2019 relig\_att\_week relig\_att\_year ///

young\_2019p crime\_2014

\* Scaling variables (min-max)

gen crime\_2019\_s = (crime\_2019 - 115.2) / (1049 - 115.2)

gen povertyp\_2019\_s = (povertyp\_2019 - 4.9) / (19.4 - 4.9)

gen income\_2019\_s = (income\_2019\_n - 48200) / (91900 - 48200)

gen unemployment\_2019\_s = (unemployment\_2019 - 2.1) / (5.6 - 2.1)

gen crime\_2014\_s = (crime\_2014 - 99.3) / (1244.4 - 99.3)

sum crime\_2019\_s povertyp\_2019\_s income\_2019\_s unemployment\_2019\_s crime\_2014\_s

\*Exploring the each independent variable relation with treat

graph box male\_2014, over(treat)

graph box young\_2014p, over(treat)

graph box middle\_aged\_2014p, over(treat)

graph box unemployment\_2014\_s, over(treat)

graph box conservative, over(treat)

graph box catholic, over(treat)

\*Logistic Regression Models

\*Treatment selection model 1

logit treat male\_2014 young\_2014p middle\_aged\_2014p ///

unemployment\_2014\_s conservative catholic liberal ///

evangelicalprotestant historicallyblackprotestant jewish ///

mainlineprotestant unaffiliated relig\_att\_week relig\_att\_year

estimates store model1

\*Propensity scores

predict ps1, p

\*tabulating jewish to see variation in observations

tab jewish

\*Correlation among treatment model variables

corr male\_2014 young\_2014p middle\_aged\_2014p ///

unemployment\_2014\_s conservative catholic liberal ///

evangelicalprotestant historicallyblackprotestant jewish ///

mainlineprotestant unaffiliated relig\_att\_week relig\_att\_year

\*roc

roctab treat ps1

estat classification

\* Generate a variable indicating the predicted treatment status based on a threshold

gen treat\_predicted1 = (ps1 > 0.5)

\* Cross-tabulate observed and predicted treatment status

tab treat treat\_predicted1, matcell(CM1)

matrix list CM1

\* Calculate type 1 error rate (false positive rate)

gen type1\_err\_m1 = treat\_predicted1 & (treat == 0)

\* Calculate type 2 error rate (false negative rate)

gen type2\_err\_m1 = !treat\_predicted1 & (treat == 1)

sum type1\_err\_m1 type2\_err\_m1

\*removing variables based on model1 output and correlation

\*Treatment selection model2

logit treat male\_2014 young\_2014p unemployment\_2014\_s ///

catholic evangelicalprotestant historicallyblackprotestant ///

mainlineprotestant unaffiliated relig\_att\_year

estimates store model2

\*Propensity scores

predict ps2, p

\*Correlation matrix for model2 variables

corr male\_2014 young\_2014p unemployment\_2014\_s ///

catholic evangelicalprotestant historicallyblackprotestant ///

mainlineprotestant unaffiliated relig\_att\_year

\*roc

roctab treat ps2

\* Generate a variable indicating the predicted treatment status based on a threshold

gen treat\_predicted2 = (ps2 > 0.5)

\* Cross-tabulate observed and predicted treatment status

tab treat treat\_predicted2, matcell(CM2)

matrix list CM2

\* Calculate type 1 error rate (false positive rate)

gen type1\_err\_m2 = treat\_predicted2 & (treat == 0)

\* Calculate type 2 error rate (false negative rate)

gen type2\_err\_m2 = !treat\_predicted2 & (treat == 1)

sum type1\_err\_m2 type2\_err\_m2

\*Treatment selection model3

logit treat male\_2014 young\_2014p unemployment\_2014\_s catholic ///

unaffiliated

estimates store model3

\*Propensity scores

predict ps3, p

\* Generate a variable indicating the predicted treatment status based on a threshold

gen treat\_predicted3 = (ps3 > 0.5)

\* Cross-tabulate observed and predicted treatment status

tab treat treat\_predicted3, matcell(CM3)

matrix list CM3

\* Calculate type 1 error rate (false positive rate)

gen type1\_err\_m3 = treat\_predicted3 & (treat == 0)

\* Calculate type 2 error rate (false negative rate)

gen type2\_err\_m3 = !treat\_predicted3 & (treat == 1)

sum type1\_err\_m3 type2\_err\_m3

\*Treatment selection model 4

logit treat male\_2014 young\_2014p unemployment\_2014\_s catholic ///

evangelicalprotestant

estimates store model4

predict ps4, p

\* Generate a variable indicating the predicted treatment status based on a threshold

gen treat\_predicted4 = (ps4 > 0.5)

\* Cross-tabulate observed and predicted treatment status

tab treat treat\_predicted4, matcell(CM4)

matrix list CM4

\* Calculate type 1 error rate (false positive rate)

gen type1\_err\_m4 = treat\_predicted4 & (treat == 0)

\* Calculate type 2 error rate (false negative rate)

gen type2\_err\_m4 = !treat\_predicted4 & (treat == 1)

sum type1\_err\_m4 type2\_err\_m4

\*selecting model3 due to auc

roctab treat ps3

roctab treat ps4

etable, estimates(model\*) cstat(\_r\_b) cstat(\_r\_se)cstat(\_r\_p) ///

mstat(r2\_p) export(logisticmodel.docx, replace)

\*Histogram

twoway (histogram ps3 if treat==1, ///

fraction color(gs8) start(0) width(.4)) ///

(histogram ps3 if treat==0, fraction fcolor(none) ///

lcolor(black) start(0) width(.4)), ///

legend(order(1 "treatment" 2 "control" )) scheme(s1mono)

\* Covariate overlap scatter plot

twoway (scatter unemployment\_2014\_s male\_2014 if treat==1, ms(th) ) ///

(scatter unemployment\_2014\_s male\_2014 if treat==0, ms(oh) ) , ///

ytitle("Unemployment rate in 2014") ///

xtitle("Percentage of Catholics") ///

title("Overlap in Two Dimensions") ///

legend(label(1 "Treatment") label(2 "Control")) ///

scheme(s1mono) saving(projscatter, replace)

\* Covariate overlap scatter plot

twoway (scatter catholic young\_2014p if treat==1, ms(th) ) ///

(scatter catholic young\_2014p if treat==0, ms(sh) ) , ///

ytitle("Percentage of Catholics") ///

xtitle("percentage of young population") ///

title("Overlap in Two Dimensions") ///

legend(label(1 "Treatment") label(2 "Control")) ///

scheme(s1mono) saving(projscatter, replace)

\*Inverse Probability weight adjusted regression

teffects ipwra (crime\_2019\_s povertyp\_2019\_s edu\_att\_2019 income\_2019\_s ///

male\_2019 unemployment\_2019\_s relig\_att\_week relig\_att\_year ///

young\_2019p crime\_2014\_s) (treat male\_2014 young\_2014p unemployment\_2014\_s ///

catholic evangelicalprotestant, logit)

\*Simpler propensity score matching

\*Inverse probabilit weights

gen ipw=1/ps3 if treat==1

replace ipw=1/(1-ps3) if treat==0

\*Regression adjusted model for control and treatment groups

regress crime\_2019\_s povertyp\_2019\_s edu\_att\_2019 income\_2019\_s ///

male\_2019 unemployment\_2019\_s relig\_att\_week relig\_att\_year ///

young\_2019p crime\_2014\_s ///

if treat==0 [aw=ipw]

predict pcontrol

regress crime\_2019\_s povertyp\_2019\_s edu\_att\_2019 income\_2019\_s ///

male\_2019 unemployment\_2019\_s relig\_att\_week relig\_att\_year ///

young\_2019p crime\_2014\_s ///

if treat==1 [aw=ipw]

predict ptreat

\*Calculate ATU, ATT, ATE

gen teffect=ptreat-pcontrol

tabulate treat, summarize(teffect)

# **Appendix B: Log file for the analysis**

----------------------------------------------------------------------------------------------------------------

name: <unnamed>

log: D:\Masters\Sem4\Econometrics\Project\Econometrics\_final\_project.log

log type: text

opened on: 10 Dec 2023, 14:11:58

.

. \*set working directory

. cd "D:\Masters\Sem4\Econometrics\Project"

D:\Masters\Sem4\Econometrics\Project

.

. import delimited "D:\Masters\Sem4\Econometrics\Project\data\Final\_data.csv" // import the data

(encoding automatically selected: ISO-8859-1)

(33 vars, 51 obs)

.

. \*Data Exploration

. \* Generate the "treat" variable

. gen treat = .

(51 missing values generated)

.

. \* Set "treat" to 0 for states with MML before 2016

. replace treat = 0 if firstyearmml < 2016

(24 real changes made)

. \* Set "treat" to 1 for states with MML is missing

. replace treat = 1 if missing(firstyearmml) & treat != 0

(19 real changes made)

.

. \* Drop states where "treat" is missing

. drop if missing(treat)

(8 observations deleted)

.

. \*Converting string variable into numerical variable

. gen crime\_2014\_no\_commas = subinstr(violent\_crimepc\_2014, ",", "", .)

. gen crime\_2014 = real(crime\_2014\_no\_commas)

. gen crime\_2019\_no\_commas = subinstr(violent\_crimespc\_2019, ",", "", .)

. gen crime\_2019 = real(crime\_2019\_no\_commas)

. gen income\_2019\_no\_commas = subinstr(income\_2019, ",", "", .)

. gen income\_2019\_n = real(income\_2019\_no\_commas)

. drop income\_2019\_no\_commas income\_2019

.

.

. \*Summary statistics for all the variables in treatment model

. summarize male\_2014 young\_2014p middle\_aged\_2014p ///

> unemployment\_2014 conservative catholic liberal catholic ///

> evangelicalprotestant historicallyblackprotestant jewish ///

> mainlineprotestant unaffiliated relig\_att\_week relig\_att\_year

Variable | Obs Mean Std. dev. Min Max

--------------------+-------------------------------------------------------------------

male\_2014 | 43 .4937225 .0084318 .4748901 .5236804

young\_2014p | 43 .24824 .0201119 .2122993 .3571403

middle~2014p | 43 .3894358 .0146753 .3595499 .4255393

unemplo~2014 | 43 6.246512 1.376218 3.5 9.2

conservative | 43 .3894358 .0146753 .3595499 .4255393

--------------------+--------------------------------------------------------------------

catholic | 43 .1944186 .0869675 .04 .42

liberal | 43 .3799701 .0118549 .35867 .4075035

catholic | 43 .1944186 .0869675 .04 .42

evangelica~t | 43 .2555814 .1113618 .08 .52

historical~t | 43 .0588372 .0604432 .01 .24

------------------+----------------------------------------------------------------------

jewish | 43 .0172093 .0136845 .01 .07

mainlinepr~t | 43 .1606977 .0532467 .1 .32

unaffiliated | 43 .2374419 .0577801 .12 .37

relig\_att\_~k | 43 .3462791 .0735178 .21 .51

relig\_att\_~r | 43 .3286047 .0322624 .24 .41

.

. \* Scaling variable (min-max)

. gen unemployment\_2014\_s = (unemployment\_2014 - 3.5) / (9.2 - 3.5)

. sum unemployment\_2014\_s

Variable | Obs Mean Std. dev. Min Max

--------------------+-----------------------------------------------------------------

unemployme~s | 43 .4818441 .2414417 0 .9999999

.

. \*Summary statistics for all the variables in outcome model

. summarize crime\_2019 povertyp\_2019 edu\_att\_2019 income\_2019\_n ///

> male\_2019 unemployment\_2019 relig\_att\_week relig\_att\_year ///

> young\_2019p crime\_2014

Variable | Obs Mean Std. dev. Min Max

-------------------+-----------------------------------------------------------------------

crime\_2019 | 43 380.8 186.7833 115.2 1049

poverty~2019 | 43 10.55116 2.72566 4.9 19.4

edu\_att\_2019 | 43 .884186 .02381 .842 .96

income\_201~n | 43 68139.53 11361.57 48200 91900

male\_2019 | 43 .4938566 .0084034 .4742635 .5213869

-------------------+-----------------------------------------------------------------------

unemplo~2019 | 43 3.562791 .8023843 2.1 5.6

relig\_att\_~k | 43 .3462791 .0735178 .21 .51

relig\_att\_~r | 43 .3286047 .0322624 .24 .41

young\_2019p | 43 .2440332 .0184983 .2119621 .3433629

crime\_2014 | 43 364.3651 188.4314 99.3 1244.4

.

. \* Scaling variables (min-max)

. gen crime\_2019\_s = (crime\_2019 - 115.2) / (1049 - 115.2)

. gen povertyp\_2019\_s = (povertyp\_2019 - 4.9) / (19.4 - 4.9)

. gen income\_2019\_s = (income\_2019\_n - 48200) / (91900 - 48200)

. gen unemployment\_2019\_s = (unemployment\_2019 - 2.1) / (5.6 - 2.1)

. gen crime\_2014\_s = (crime\_2014 - 99.3) / (1244.4 - 99.3)

. sum crime\_2019\_s povertyp\_2019\_s income\_2019\_s unemployment\_2019\_s crime\_2014\_s

Variable | Obs Mean Std. dev. Min Max

-------------------+-----------------------------------------------------------------

crime\_2019\_s | 43 .2844292 .200025 -3.27e-09 1

povertyp\_2~s | 43 .3897354 .1879766 6.58e-09 1

income\_201~s | 43 .4562823 .2599902 0 1

unemploy~9\_s | 43 .4179402 .2292527 -2.72e-08 1

crime\_2014\_s | 43 .2314777 .1645545 2.67e-09 1

.

. \*Exploring the each independent variable relation with treat

. graph box male\_2014, over(treat)

. graph box young\_2014p, over(treat)

. graph box middle\_aged\_2014p, over(treat)

. graph box unemployment\_2014\_s, over(treat)

. graph box conservative, over(treat)

. graph box catholic, over(treat)

.

. \*Logistic Regression Models

. \*Treatment selection model 1

. logit treat male\_2014 young\_2014p middle\_aged\_2014p ///

> unemployment\_2014\_s conservative catholic liberal ///

> evangelicalprotestant historicallyblackprotestant jewish ///

> mainlineprotestant unaffiliated relig\_att\_week relig\_att\_year

note: jewish != .01 predicts failure perfectly;

jewish omitted and 15 obs not used.

note: conservative omitted because of collinearity.

Iteration 0: Log likelihood = -17.582364

Iteration 1: Log likelihood = -2.7749057

Iteration 2: Log likelihood = -.27614191

Iteration 3: Log likelihood = 0

Iteration 4: Log likelihood = 0

Logistic regression Number of obs = 28

LR chi2(-1) = 35.16

Prob > chi2 = .

Log likelihood = 0 Pseudo R2 = 1.0000

-------------------------------------------------------------------------------------------------

treat | Coefficient Std. err. z P>|z| [95% conf. interval]

--------------------------------+----------------------------------------------------------------

male\_2014 | 5196.888 . . . . .

young\_2014p | -10566.11 . . . . .

middle\_aged\_2014p | -21238.26 . . . . .

unemployment\_2014\_s | -110.8234 . . . . .

conservative | 0 (omitted)

catholic | 866.2159 . . . . .

liberal | 21252.76 . . . . .

evangelicalprotestant | 746.4045 . . . . .

historicallyblackprotestant | 783.2877 . . . . .

jewish | 0 (omitted)

mainlineprotestant | 270.1072 . . . . .

unaffiliated | 2154.084 . . . . .

relig\_att\_week | 1999.557 . . . . .

relig\_att\_year | 1295.605 . . . . .

\_cons | -1831.219 . . . . .

---------------------------------------------------------------------------------------------

Note: 9 failures and 19 successes completely determined.

. estimates store model1

.

. \*Propensity scores

. predict ps1, p

(15 missing values generated)

.

. \*tabulating jewish to see variation in observations

. tab jewish

Jewish | Freq. Percent Cum.

------------+-----------------------------------

.01 | 28 65.12 65.12

.02 | 8 18.60 83.72

.03 | 4 9.30 93.02

.05 | 1 2.33 95.35

.06 | 1 2.33 97.67

.07 | 1 2.33 100.00

-------------+-----------------------------------

Total | 43 100.00

.

. \*Correlation among treatment model variables

. corr male\_2014 young\_2014p middle\_aged\_2014p ///

> unemployment\_2014\_s conservative catholic liberal ///

> evangelicalprotestant historicallyblackprotestant jewish ///

> mainlineprotestant unaffiliated relig\_att\_week relig\_att\_year

(obs=43)

| mal~2014 yo~2014p mi~2014p unem~4\_s conser~e catholic liberal evange~t histor~t jewish mainli~t unaffi~d relig\_~k relig\_~r

-------------+------------------------------------------------------------------------------------------------------------------------------

male\_2014 | 1.0000

young\_2014p | -0.0900 1.0000

middle~2014p | -0.2032 -0.5340 1.0000

unemploy~4\_s | -0.3677 0.1629 0.1403 1.0000

conservative | -0.2032 -0.5340 1.0000 0.1403 1.0000

catholic | -0.0745 0.0163 0.2250 0.2101 0.2250 1.0000

liberal | -0.2797 -0.4573 0.9673 0.2558 0.9673 0.2424 1.0000

evangelica~t | 0.0475 -0.1600 -0.2888 -0.0227 -0.2888 -0.7623 -0.2906 1.0000

historical~t | -0.6143 0.3631 -0.0870 0.2915 -0.0870 -0.4298 0.0255 0.2627 1.0000

jewish | -0.4433 0.2444 0.2101 0.2422 0.2101 0.4487 0.2926 -0.5817 0.1601 1.0000

mainlinepr~t | 0.1324 -0.2923 -0.0779 -0.6272 -0.0779 -0.0243 -0.1755 -0.0424 -0.2313 -0.2717 1.0000

unaffiliated | 0.3336 -0.0061 0.3712 -0.0685 0.3712 0.2766 0.2775 -0.5927 -0.5026 0.1714 -0.1828 1.0000

relig\_att\_~k | -0.1564 -0.0192 -0.4099 0.0427 -0.4099 -0.5545 -0.3451 0.8047 0.4737 -0.3703 0.0116 -0.8615 1.0000

relig\_att\_~r | 0.0226 0.0408 -0.0370 -0.0092 -0.0370 0.1923 0.0213 -0.1807 0.1054 0.2013 0.1420 -0.1207 -0.1980 1.0000

.

. \*roc

. roctab treat ps1

ROC Asymptotic normal

Obs area Std. err. [95% conf. interval]

------------------------------------------------------------

28 1.0000 0.0000 1.00000 1.00000

. estat classification

Logistic model for treat

-------- True --------

Classified | D ~D | Total

--------------+---------------------------+-----------

+ | 19 0 | 19

- | 0 9 | 9

--------------+--------------------------+-----------

Total | 19 9 | 28

Classified + if predicted Pr(D) >= .5

True D defined as treat != 0

--------------------------------------------------------

Sensitivity Pr( +| D) 100.00%

Specificity Pr( -|~D) 100.00%

Positive predictive value Pr( D| +) 100.00%

Negative predictive value Pr(~D| -) 100.00%

----------------------------------------------------------

False + rate for true ~D Pr( +|~D) 0.00%

False - rate for true D Pr( -| D) 0.00%

False + rate for classified + Pr(~D| +) 0.00%

False - rate for classified - Pr( D| -) 0.00%

----------------------------------------------------------

Correctly classified 100.00%

----------------------------------------------------------

. \* Generate a variable indicating the predicted treatment status based on a threshold

. gen treat\_predicted1 = (ps1 > 0.5)

. \* Cross-tabulate observed and predicted treatment status

. tab treat treat\_predicted1, matcell(CM1)

| treat\_predicted1

treat | 0 1 | Total

-----------+----------------------+----------

0 | 9 15 | 24

1 | 0 19 | 19

-----------+----------------------+----------

Total | 9 34 | 43

. matrix list CM1

CM1[2,2]

c1 c2

r1 9 15

r2 0 19

. \* Calculate type 1 error rate (false positive rate)

. gen type1\_err\_m1 = treat\_predicted1 & (treat == 0)

. \* Calculate type 2 error rate (false negative rate)

. gen type2\_err\_m1 = !treat\_predicted1 & (treat == 1)

. sum type1\_err\_m1 type2\_err\_m1

Variable | Obs Mean Std. dev. Min Max

-----------------+---------------------------------------------------------

type1\_err\_m1 | 43 .3488372 .4822428 0 1

type2\_err\_m1 | 43 0 0 0 0

. \*removing variables based on model1 output and correlation

. \*Treatment selection model2

. logit treat male\_2014 young\_2014p unemployment\_2014\_s ///

> catholic evangelicalprotestant historicallyblackprotestant ///

> mainlineprotestant unaffiliated relig\_att\_year

Iteration 0: Log likelihood = -29.513972

Iteration 1: Log likelihood = -8.6834114

Iteration 2: Log likelihood = -7.6118725

Iteration 3: Log likelihood = -6.8411937

Iteration 4: Log likelihood = -6.8086234

Iteration 5: Log likelihood = -6.8081612

Iteration 6: Log likelihood = -6.8081606

Logistic regression Number of obs = 43

LR chi2(9) = 45.41

Prob > chi2 = 0.0000

Log likelihood = -6.8081606 Pseudo R2 = 0.7693

---------------------------------------------------------------------------------------------------------------

treat | Coefficient Std. err. z P>|z| [95% conf. interval]

-----------------------------------+---------------------------------------------------------------------------

male\_2014 | 120.5773 277.5798 0.43 0.664 -423.4692 664.6238

young\_2014p | -147.4405 181.3358 -0.81 0.416 -502.8521 207.9711

unemployment\_2014\_s | -3.034611 5.888808 -0.52 0.606 -14.57646 8.507241

catholic | -27.01224 26.37493 -1.02 0.306 -78.70616 24.68168

evangelicalprotestant | 38.03175 22.62283 1.68 0.093 -6.308193 82.37168

historicallyblackprotestant | -5.018355 40.42856 -0.12 0.901 -84.25689 74.22018

mainlineprotestant | -8.079001 22.67091 -0.36 0.722 -52.51316 36.35516

unaffiliated | -73.15237 54.06383 -1.35 0.176 -179.1155 32.81078

relig\_att\_year | 3.630141 27.95332 0.13 0.897 -51.15735 58.41764

\_cons | -8.993147 146.2706 -0.06 0.951 -295.6782 277.6919

----------------------------------------------------------------------------------------------------------------

Note: 2 failures and 2 successes completely determined.

. estimates store model2

. \*Propensity scores

. predict ps2, p

.

. \*Correlation matrix for model2 variables

. corr male\_2014 young\_2014p unemployment\_2014\_s ///

> catholic evangelicalprotestant historicallyblackprotestant ///

> mainlineprotestant unaffiliated relig\_att\_year

(obs=43)

| mal~2014 yo~2014p unem~4\_s catholic evange~t histor~t mainli~t unaffi~d relig\_~r

------------------+---------------------------------------------------------------------------------

male\_2014 | 1.0000

young\_2014p | -0.0900 1.0000

unemploy~4\_s | -0.3677 0.1629 1.0000

catholic | -0.0745 0.0163 0.2101 1.0000

evangelica~t | 0.0475 -0.1600 -0.0227 -0.7623 1.0000

historical~t | -0.6143 0.3631 0.2915 -0.4298 0.2627 1.0000

mainlinepr~t | 0.1324 -0.2923 -0.6272 -0.0243 -0.0424 -0.2313 1.0000

unaffiliated | 0.3336 -0.0061 -0.0685 0.2766 -0.5927 -0.5026 -0.1828 1.0000

relig\_att\_~r | 0.0226 0.0408 -0.0092 0.1923 -0.1807 0.1054 0.1420 -0.1207 1.0000

.

. \*roc

. roctab treat ps2

ROC Asymptotic normal

Obs area Std. err. [95% conf. interval]

------------------------------------------------------------

43 0.9868 0.0124 0.96244 1.00000

.

. \* Generate a variable indicating the predicted treatment status based on a threshold

. gen treat\_predicted2 = (ps2 > 0.5)

. \* Cross-tabulate observed and predicted treatment status

. tab treat treat\_predicted2, matcell(CM2)

| treat\_predicted2

treat | 0 1 | Total

------------+----------------------+----------

0 | 24 0 | 24

1 | 2 17 | 19

------------+----------------------+----------

Total | 26 17 | 43

. matrix list CM2

CM2[2,2]

c1 c2

r1 24 0

r2 2 17

. \* Calculate type 1 error rate (false positive rate)

. gen type1\_err\_m2 = treat\_predicted2 & (treat == 0)

. \* Calculate type 2 error rate (false negative rate)

. gen type2\_err\_m2 = !treat\_predicted2 & (treat == 1)

. sum type1\_err\_m2 type2\_err\_m2

Variable | Obs Mean Std. dev. Min Max

-----------------+-----------------------------------------------------------

type1\_err\_m2 | 43 0 0 0 0

type2\_err\_m2 | 43 .0465116 .2130826 0 1

. \*Treatment selection model3

. logit treat male\_2014 young\_2014p unemployment\_2014\_s catholic ///

> unaffiliated

Iteration 0: Log likelihood = -29.513972

Iteration 1: Log likelihood = -11.04314

Iteration 2: Log likelihood = -9.8806921

Iteration 3: Log likelihood = -9.6850573

Iteration 4: Log likelihood = -9.6839693

Iteration 5: Log likelihood = -9.6839691

Logistic regression Number of obs = 43

LR chi2(5) = 39.66

Prob > chi2 = 0.0000

Log likelihood = -9.6839691 Pseudo R2 = 0.6719

----------------------------------------------------------------------------------------------------------

treat | Coefficient Std. err. z P>|z| [95% conf. interval]

--------------------+------------------------------------------------------------------------------------

male\_2014 | 145.6758 127.4929 1.14 0.253 -104.2056 395.5573

young\_2014p | -62.20073 81.60458 -0.76 0.446 -222.1428 97.7413

unemployment\_2014\_s | -1.407646 3.285937 -0.43 0.668 -7.847963 5.032672

catholic | -29.55417 12.63997 -2.34 0.019 -54.32806 -4.780281

unaffiliated | -57.8759 23.37021 -2.48 0.013 -103.6807 -12.07113

\_cons | -37.21666 47.37682 -0.79 0.432 -130.0735 55.6402

--------------------------------------------------------------------------------------------------------

. estimates store model3

. \*Propensity scores

. predict ps3, p

.

. \* Generate a variable indicating the predicted treatment status based on a threshold

. gen treat\_predicted3 = (ps3 > 0.5)

. \* Cross-tabulate observed and predicted treatment status

. tab treat treat\_predicted3, matcell(CM3)

| treat\_predicted3

treat | 0 1 | Total

-----------+----------------------+----------

0 | 23 1 | 24

1 | 2 17 | 19

-----------+----------------------+----------

Total | 25 18 | 43

. matrix list CM3

CM3[2,2]

c1 c2

r1 23 1

r2 2 17

. \* Calculate type 1 error rate (false positive rate)

. gen type1\_err\_m3 = treat\_predicted3 & (treat == 0)

. \* Calculate type 2 error rate (false negative rate)

. gen type2\_err\_m3 = !treat\_predicted3 & (treat == 1)

. sum type1\_err\_m3 type2\_err\_m3

Variable | Obs Mean Std. dev. Min Max

-----------------+---------------------------------------------------------

type1\_err\_m3 | 43 .0232558 .1524986 0 1

type2\_err\_m3 | 43 .0465116 .2130826 0 1

. \*Treatment selection model 4

. logit treat male\_2014 young\_2014p unemployment\_2014\_s catholic ///

> evangelicalprotestant

Iteration 0: Log likelihood = -29.513972

Iteration 1: Log likelihood = -9.8806206

Iteration 2: Log likelihood = -9.2394995

Iteration 3: Log likelihood = -9.1569818

Iteration 4: Log likelihood = -9.1567945

Iteration 5: Log likelihood = -9.1567945

Logistic regression Number of obs = 43

LR chi2(5) = 40.71

Prob > chi2 = 0.0000

Log likelihood = -9.1567945 Pseudo R2 = 0.6897

----------------------------------------------------------------------------------------------------------

treat | Coefficient Std. err. z P>|z| [95% conf. interval]

----------------------+-----------------------------------------------------------------------------------

male\_2014 | -116.7122 119.7154 -0.97 0.330 -351.3501 117.9257

young\_2014p | -12.32165 88.15417 -0.14 0.889 -185.1006 160.4573

unemployment\_2014\_s | -9.155148 4.319325 -2.12 0.034 -17.62087 -.6894256

catholic | -5.177936 13.9166 -0.37 0.710 -32.45398 22.09811

evangelicalprotestant | 36.66235 13.97475 2.62 0.009 9.272338 64.05236

\_cons | 55.93841 59.31222 0.94 0.346 -60.31141 172.1882

--------------------------------------------------------------------------------------------------------------

. estimates store model4

. predict ps4, p

.

. \* Generate a variable indicating the predicted treatment status based on a threshold

. gen treat\_predicted4 = (ps4 > 0.5)

. \* Cross-tabulate observed and predicted treatment status

. tab treat treat\_predicted4, matcell(CM4)

| treat\_predicted4

Treat | 0 1 | Total

-----------+----------------------+----------

0 | 22 2 | 24

1 | 2 17 | 19

-----------+----------------------+----------

Total | 24 19 | 43

. matrix list CM4

symmetric CM4[2,2]

c1 c2

r1 22

r2 2 17

. \* Calculate type 1 error rate (false positive rate)

. gen type1\_err\_m4 = treat\_predicted4 & (treat == 0)

. \* Calculate type 2 error rate (false negative rate)

. gen type2\_err\_m4 = !treat\_predicted4 & (treat == 1)

. sum type1\_err\_m4 type2\_err\_m4

Variable | Obs Mean Std. dev. Min Max

-----------------+---------------------------------------------------------

type1\_err\_m4 | 43 .0465116 .2130826 0 1

type2\_err\_m4 | 43 .0465116 .2130826 0 1

. // -----------------------------------------------------------------------------

. \*selecting model3 due to auc

. roctab treat ps3

ROC Asymptotic normal

Obs area Std. err. [95% conf. interval]

------------------------------------------------------------

43 0.9627 0.0249 0.91387 1.00000

. roctab treat ps4

ROC Asymptotic normal

Obs area Std. err. [95% conf. interval]

------------------------------------------------------------

43 0.9737 0.0204 0.93364 1.00000

.

. etable, estimates(model\*) cstat(\_r\_b) cstat(\_r\_se)cstat(\_r\_p) ///

> mstat(r2\_p) export(logisticmodel.docx, replace)

------------------------------------------------------------------------------------------------------

treat treat treat treat

-------------------------------------------------------------------------------------------------------

Male\_2014 5196.888 120.577 145.676 -116.712

(0.000) (277.580) (127.493) (119.715)

0.66 0.25 0.33

Young\_2014p -1.06e+04 -147.441 -62.201 -12.322

(0.000) (181.336) (81.605) (88.154)

0.42 0.45 0.89

Middle\_aged\_2014p -2.12e+04

(0.000)

unemployment\_2014\_s -110.823 -3.035 -1.408 -9.155

(0.000) (5.889) (3.286) (4.319)

0.61 0.67 0.03

Catholic 866.216 -27.012 -29.554 -5.178

(0.000) (26.375) (12.640) (13.917)

0.31 0.02 0.71

Liberal 21252.758

(0.000)

Evangelical Protestant 746.405 38.032 36.662

(0.000) (22.623) (13.975)

0.09 0.01

Historically Black Protestant 783.288 -5.018

(0.000) (40.429)

0.90

Mainline Protestant 270.107 -8.079

(0.000) (22.671)

0.72

Unaffiliated 2154.084 -73.152 -57.876

(0.000) (54.064) (23.370)

0.18 0.01

relig\_att\_week 1999.557

(0.000)

relig\_att\_year 1295.605 3.630

(0.000) (27.953)

0.90

Intercept -1831.219 -8.993 -37.217 55.938

(0.000) (146.271) (47.377) (59.312)

0.95 0.43 0.35

Pseudo R-squared 1.00 0.77 0.67 0.69

----------------------------------------------------------------------------------------------------

(collection ETable exported to file logisticmodel.docx)

.

. \*Histogram

. twoway (histogram ps3 if treat==1, ///

> fraction color(gs8) start(0) width(.4)) ///

> (histogram ps3 if treat==0, fraction fcolor(none) ///

> lcolor(black) start(0) width(.4)), ///

> legend(order(1 "treatment" 2 "control" )) scheme(s1mono)

.

. \* Covariate overlap scatter plot

. twoway (scatter unemployment\_2014\_s male\_2014 if treat==1, ms(th) ) ///

> (scatter unemployment\_2014\_s male\_2014 if treat==0, ms(oh) ) , ///

> ytitle("Unemployment rate in 2014") ///

> xtitle("Percentage of Catholics") ///

> title("Overlap in Two Dimensions") ///

> legend(label(1 "Treatment") label(2 "Control")) ///

> scheme(s1mono) saving(projscatter, replace)

file projscatter.gph saved

.

.

. \* Covariate overlap scatter plot

. twoway (scatter catholic young\_2014p if treat==1, ms(th) ) ///

> (scatter catholic young\_2014p if treat==0, ms(sh) ) , ///

> ytitle("Percentage of Catholics") ///

> xtitle("percentage of young population") ///

> title("Overlap in Two Dimensions") ///

> legend(label(1 "Treatment") label(2 "Control")) ///

> scheme(s1mono) saving(projscatter, replace)

file projscatter.gph saved

.

. \*Inverse Probability weight adjusted regression

. teffects ipwra (crime\_2019\_s povertyp\_2019\_s edu\_att\_2019 income\_2019\_s ///

> male\_2019 unemployment\_2019\_s relig\_att\_week relig\_att\_year ///

> young\_2019p crime\_2014\_s) (treat male\_2014 young\_2014p unemployment\_2014\_s ///

> catholic evangelicalprotestant, logit)

Iteration 0 : EE criterion = 2.239e-23

Iteration 1 : EE criterion = 1.046e-31

Treatment-effects estimation Number of obs = 43

Estimator : IPW regression adjustment

Outcome model : linear

Treatment model : logit

---------------------------------------------------------------------------------------------------------

| Robust

crime\_2019\_s | Coefficient std. err. z P>|z| [95% conf. interval]

--------------------------+-----------------------------------------------------------------------------

ATE |

treat |

(1 vs 0) | -.1105853 .0288957 -3.83 0.000 -.1672198 -.0539509

--------------------------+----------------------------------------------------------------

POmean |

treat |

0 | .3211547 .0388014 8.28 0.000 .2451054 .3972041

-------------------------------------------------------------------------------------------------------

. \*Simpler propensity score matching

. \*Inverse probabilit weights

. gen ipw=1/ps3 if treat==1

(24 missing values generated)

. replace ipw=1/(1-ps3) if treat==0

(24 real changes made)

.

. \*Regression adjusted model for control and treatment groups

. regress crime\_2019\_s povertyp\_2019\_s edu\_att\_2019 income\_2019\_s ///

> male\_2019 unemployment\_2019\_s relig\_att\_week relig\_att\_year ///

> young\_2019p crime\_2014\_s ///

> if treat==0 [aw=ipw]

(sum of wgt is 32.33397042751312)

Source | SS df MS Number of obs = 24

----------------------+---------------------------------- F(9, 14) = 24.63

Model | 1.12190724 9 .12465636 Prob > F = 0.0000

Residual | .070847952 14 .005060568 R-squared = 0.9406

----------------------+----------------------------------------- Adj R-squared = 0.9024

Total | 1.19275519 23 .051858921 Root MSE = .07114

-----------------------------------------------------------------------------------------------------------

crime\_2019\_s | Coefficient Std. err. t P>|t| [95% conf. interval]

----------------------+-------------------------------------------------------------------------------------

povertyp\_2019\_s | .3938756 .1884834 2.09 0.055 -.0103811 .7981323

edu\_att\_2019 | -.1191554 1.263101 -0.09 0.926 -2.828239 2.589928

income\_2019\_s | .1332166 .1315721 1.01 0.328 -.1489776 .4154108

male\_2019 | 5.690938 2.177184 2.61 0.020 1.021344 10.36053

unemployment\_20| .0154634 .1552875 0.10 0.922 -.3175952 .348522

relig\_att\_week | .5382531 .5241451 1.03 0.322 -.5859264 1.662432

relig\_att\_year | -.8784622 .7609937 -1.15 0.268 -2.510631 .7537068

young\_2019p| -.8469072 1.579444 -0.54 0.600 -4.234478 2.540664

crime\_2014\_s| 1.003192 .1815629 5.53 0.000 .6137787 1.392606

\_cons | -2.525801 1.707429 -1.48 0.161 -6.187872 1.136269

---------------------------------------------------------------------------------------------------------

. predict pcontrol

(option xb assumed; fitted values)

. regress crime\_2019\_s povertyp\_2019\_s edu\_att\_2019 income\_2019\_s ///

> male\_2019 unemployment\_2019\_s relig\_att\_week relig\_att\_year ///

> young\_2019p crime\_2014\_s ///

> if treat==1 [aw=ipw]

(sum of wgt is 34.62190210819244)

Source | SS df MS Number of obs= 19

-----------------------------------+---------------------------------------- F(9, 9) = 36.98

Model | .173585241 9 .019287249 Prob > F = 0.0000

Residual | .004694605 9 .000521623 R-squared = 0.9737

-----------------------------------+---------------------------------------- Adj R-squared = 0.9473

Total | .178279846 18 .009904436 Root MSE = .02284

----------------------------------------------------------------------------------------------------------------

crime\_2019\_s | Coefficient Std. err. t P>|t| [95% conf. interval]

-----------------------------------+----------------------------------------------------------------------------

povertyp\_2019\_s | -.2412215 .1459962 -1.65 0.133 -.5714879 .0890449

edu\_att\_2019 | -.0339941 .5364872 -0.06 0.951 -1.247612 1.179624

income\_2019\_s | -.2538557 .1034269 -2.45 0.036 -.4878236 -.0198878

male\_2019 | 1.691939 1.681574 1.01 0.341 -2.112045 5.495922

unemployment\_2019\_s | -.0208448 .0747226 -0.28 0.787 -.1898791 .1481895

relig\_att\_week | .8314395 .2865653 2.90 0.018 .1831837 1.479695

relig\_att\_year | .7720672 .2666942 2.89 0.018 .168763 1.375371

young\_2019p | .4657507 1.50691 0.31 0.764 -2.943117 3.874618

crime\_2014\_s | 1.081997 .1206847 8.97 0.000 .8089889 1.355004

\_cons | -1.279063 1.201666 -1.06 0.315 -3.997419 1.439294

----------------------------------------------------------------------------------------------------------------

. predict ptreat

(option xb assumed; fitted values)

.

. \*Calculate ATU, ATT, ATE

. gen teffect=ptreat-pcontrol

. tabulate treat, summarize(teffect)

| Summary of teffect

treat | Mean Std. dev. Freq.

-----------------+------------------------------------------

0 | -.1302025 .08127583 24

1 | -.08388012 .08785628 19

-----------------+--------------------------------------------

Total | -.10973447 . 08641327 43

.

end of do-file

# **References**

1. [Jacob Felson, Amy Adamczyk, Christopher Thomas(2019) How and why have attitudes about cannabis legalization changed so much?](1.%09https:/www.sciencedirect.com/science/article/abs/pii/S0049089X17310232" \l ":~:text=or%20religious%20engagement.-,Changes%20in%20cannabis%20use%20have%20had%20minimal%20effects%20on%20attitudes,framing%20all%20played%20contributing%20roles.)

1. [Miech, R. A., Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2015). Trends in U.S., past-year marijuana use from 1985 to 2009: An age–period–cohort analysis.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4109063/" \l ":~:text=The%20three%20most%20recent%2C%20quinquennial,increase%20from%202000%20to%202009.)

1. [Pacula, R. L., Powell, D., Heaton, P., & Sevigny, E. L. (2015). Social and Political Factors Associated With State-Level Legalization of Cannabis in the United States.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8022331/)

1. [Dragone, D., Prarolo, G., Vanin, P., & Zanella, G. (2019). Crime and the legalization of recreational marijuana.](https://www.sciencedirect.com/science/article/abs/pii/S0167268118300386" \l ":~:text=We%20find%20across%20different%20specifications,and%20100%20per%20100%2C000%20inhabitants).)

1. [Burdette, A. M., Haynes, C., & Higgins, M. (2018). Religious Involvement and Marijuana Use for Medical and Recreational Purposes.](https://journals.sagepub.com/doi/full/10.1177/0022042618770393" \l "table2-0022042618770393)
2. Pew Research Center. (n.d.). Political ideology distribution in the U.S. Retrieved from[ https://www.pewresearch.org/religion/religious-landscape-study/compare/political-ideology/by/state/ ]
3. Pew Research Center. (n.d.). Religious Affiliation data. Retrieved from [ https://www.pewresearch.org/religion/religious-landscape-study/compare/religious-tradition/by/state/]
4. Bureau of Labor Statistics. (n.d.). Unemployment rate. Retrieved from [https://www.bls.gov/opub/ted/2014/ted\_20140319.htm]
5. U.S. Census Bureau. (n.d.). Age distribution. Retrieved from [https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html]
6. U.S. Census Bureau. (n.d.). Gender Distribution. Retrieved from [https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html]
7. National Center for Education Statistics. (n.d.). Percentage of Education Attainment. Retrieved from [https://nces.ed.gov/programs/digest/d95/dtab011.asp]
8. U.S. Census Bureau. (n.d.). Average state Poverty rate. Retrieved from [https://www.census.gov/library/publications/2020/demo/p60-270.html]
9. Federal Bureau of Investigation - Uniform Crime Reporting. (n.d.). Violent Crime Rates. Retrieved from [https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/tables/table-4]