**Project Draft Analysis**

# **Introduction:**

The intersection of drug policy and its impact on societal well-being has been a subject of growing interest and scrutiny. In recent years, the liberalization of marijuana laws in various U.S. states has ignited discussions about its potential effects on crime rates, particularly violent crimes. This study delves into the complex correlation between marijuana law liberalization and violent crimes, seeking to provide empirical insights into this multifaceted issue.

Marijuana, once widely criminalized, has experienced a transformative shift in its legal status across the United States. As of 2019, 31 states and the District of Columbia have legalized marijuana for either medicinal or recreational use, reflecting a broader trend toward relaxation of restrictions. This changing landscape prompts crucial questions regarding the potential consequences of marijuana law liberalization on public safety, specifically in relation to violent crimes.

The primary objective of this project is to analyse the impact of marijuana law liberalization on violent crime rates at the state level. By employing robust statistical methods, including the application of propensity scores that contribute to uncover patterns, associations, and potential causal relationships that contribute to a nuanced understanding of the issue. Through the investigation of treatment and outcome models, we seek to elucidate the intricate dynamics between marijuana policy changes and violent crime trends.

This project holds significance for religious affiliation, unemployment rate, demographic data such as age distribution and gender distribution grappling with the implications of evolving marijuana legislation. Furthermore, this project sheds light on the interplay between marijuana law liberalization and violent crimes and aims to contribute evidence-based insights that can guide informed decision-making.

# **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, we 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:**

Our 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 (young\_2014p, middle\_aged\_2014p)

Religious affiliation (catholic)

Gender distribution (male\_2014)

Unemployment rate (unemployment\_2014)

Unobserved

Marijuana Legalization

Political ideology (conservative)

1. Gender distribution in 2014(male\_2014): The inclusion of the percentage of the male population in 2014 considers societal gender dynamics, which might play a role in shaping attitudes towards marijuana liberalization.
2. Age distribution (young\_2014p, middle\_aged\_2014p): The percentage of young people and middle aged people in 2014 is included, recognizing that younger demographics might have different perspectives on marijuana policies compared to older age groups.
3. Unemployment rate (unemployment\_2014): The unemployment rate in 2014 is a crucial economic factor that could affect a state's stance on marijuana liberalization, as economic conditions often shape policy decisions.
4. Political ideology(conservative): The percentage of conservatives in the population of 2014 is included to capture the ideological landscape within a state, as conservative views may influence the adoption of marijuana liberalization.
5. Religion (catholic): The percentage of Catholics in the population of 2014 is considered, acknowledging that religious affiliations can impact societal views and potentially influence policy decisions.

## **Exploratory analysis using Box Plots:**

In an effort to understand the distributional characteristics of key variables and their association with the treatment variable, a series of box plots were generated. These box plots provide a visual representation of the variation in each variable between the treated and control groups, offering insights into potential imbalances and aiding in the assessment of covariate balance.

**A diagram of a graph

Description automatically generated with medium confidence1. Male and young percentage in 2014:**

This box plots explores the distribution of the percentage of male population in 2014 (male\_2014) and percentage of young population (Young\_2014) between the treated and control groups. The y-axis represents the variable's values, while the x-axis distinguishes between the two groups. it is evident from the box plot that there is a notable similarity in the distribution between the treatment and control groups.

A screenshot of a graph

Description automatically generatedThis visual evidence suggests that, with regard to the variables male\_2014 and young\_2014, there is a degree of covariate balance achieved between the treated and control groups. The proportion of male and young population in 2014 does not exhibit substantial differences, indicating a potential successful initial balancing of this particular covariate.

**A diagram of a graph

Description automatically generated with medium confidence2. Middle aged population:**

The box plot reveals that the median and interquartile range of the middle-aged population percentage are noticeably higher in the control group compared to the treatment group. Additionally, the whiskers of the box plot, representing the variability outside the upper and lower quartiles, show an expanded spread of middle-aged population proportions in the control group.

This visual evidence suggests an existing imbalance in the distribution of the variable “middle\_aged\_2014p” between the treated and control groups. The proportion of middle-aged population in 2014 exhibits a notable disparity, indicating a potential need for further adjustments or consideration during subsequent stages of the analysis.

3. **Unemployment** **rate**:

A screenshot of a graph

Description automatically generatedThe box plot reveals variations in the median and interquartile range of the unemployment rate between the treatment and control groups. Notably, the median unemployment rate in the treatment group appears slightly higher than that in the control group.

This visual evidence indicates a discernible difference in the distribution of the variable "unemployment\_2014" between the treated and control groups. Understanding these disparities is crucial for assessing the covariate balance, as imbalances in certain covariates may impact the validity of subsequent causal inferences drawn from the analysis.

**A screenshot of a graph

Description automatically generated4. Conservative**:

The box plot reveals variations in the median and interquartile range of the conservative between the treatment and control groups. Notably, the median conservative in the treatment group appears slightly higher than that in the control group



1. **Catholic:**

The box plot indicates a degree of imbalance, suggesting that the percentage of catholic population in 2014 may vary between the treatment and control groups.

While certain variables exhibit balanced distributions, others reveal imbalances that warrant careful consideration during subsequent analyses. These findings underscore the importance of further statistical adjustments to enhance covariate balance, ensuring the validity of causal inferences drawn from the matching model.

## **Results of logistic regression: Treatment model:**

Logistic regression Model results

|  |  |
| --- | --- |
|  | treat |
| Male\_2014 | -109.463 |
|  | (89.223) |
|  | 0.22 |
| Young\_2014p | -59.118 |
|  | (37.940) |
|  | 0.12 |
| Middle\_aged\_2014p | -145.221 |
|  | (62.647) |
|  | 0.02 |
| Unemployment\_2014 | -0.912 |
|  | (0.535) |
|  | 0.09 |
| Catholic | -32.826 |
|  | (12.966) |
|  | 0.01 |
| Intercept | 136.214 |
|  | (66.849) |
|  | 0.04 |
| Pseudo R-squared | 0.57 |

Note: In above table, 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.

The logistic regression analysis aimed to model the likelihood of a state adopting marijuana liberalization, serving as the treatment variable. The results offer insights into the estimated coefficients, significance levels, and odds ratios for each covariate. The percentage of males in 2014 did not significantly influence the likelihood of marijuana liberalization (Coefficient: -109.463, p = 0.220). Similarly, the percentage of young population in 2014 showed a non-significant association with the likelihood of liberalization (Coefficient: -59.11765, p = 0.119). In contrast, a higher percentage of middle-aged population in 2014 was associated with a lower likelihood of liberalization and was statistically significant (Coefficient: -145.2213, p = 0.020). The unemployment rate in 2014 exhibited a marginally significant negative association (Coefficient: -0.9122639, p = 0.088), suggesting that higher unemployment rates were linked to a lower likelihood of liberalization. Notably, a higher percentage of Catholics in 2014 was statistically significant in predicting a lower likelihood of liberalization (Coefficient: -32.82592, p = 0.011). It is important to note that the variable "conservative" was omitted due to collinearity. These findings underscore the intricate relationship between demographic and socio-economic factors and state-level marijuana policy decisions.

Note: The variable "conservative" was omitted due to collinearity.

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

A graph of a patient

Description automatically generated with medium confidenceTo 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.

## **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.

A graph of a number of people

Description automatically generated with medium confidence

**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 male and young population in 2019, the unemployment rate in 2019, and the baseline violent crime rates in 2014. These covariates are strategically chosen to control for potential confounding factors that might influence the relationship between marijuana policy changes and crime rates.

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)

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 is linked to the outcome, emphasizing the importance of age distribution 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.

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 (IPWRA) 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. For the control group, where the treatment variable was 0, the results revealed significant impacts of covariates on the outcome variable, crime rates in 2019. Notably, variables such as male population in 2019, unemployment rate in 2019, percentage of young population in 2019, and crime rates in 2014 showed substantial influence. The model exhibited a good fit with an R-squared of 0.8677 and an adjusted R-squared of 0.8382.

A parallel analysis for the treatment group (treatment variable = 1) demonstrated the effects of the same covariates, indicating the potential impact of marijuana legalization on crime rates. The regression for the treatment group exhibited a higher R-squared of 0.9395 and an adjusted R-squared of 0.9222, suggesting a stronger fit.

Upon calculating the treatment effect using the propensity score matching method, the analysis revealed interesting insights into the impact of marijuana legalization on violent crime rates. The mean treatment effect across all states was found to be -8.38, indicating a general reduction in violent crimes associated with marijuana liberalization. However, it's crucial to note the considerable variation in the treatment effect, as reflected by the standard deviation of 45.87. This variability underscores the complexity of the relationship between marijuana policy changes and crime rates, suggesting that the effect is not uniformly experienced across all states. Further examination by treatment status showed that, on average, states with marijuana legalization experienced a smaller reduction in crime rates (-1.27) compared to states without legalization (-14.26). These findings highlight the nuanced nature of the treatment effect, emphasizing the need for careful consideration of contextual factors influencing the impact of marijuana law liberalization on violent crimes.

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

The results obtained from the inverse probability weight adjusted regression analysis do not provide clear evidence of a causal effect of marijuana legalization on violent crimes. The Average Treatment Effect (ATE) represents the estimated difference in violent crime rates between states with and without marijuana law liberalization. The coefficient for the “treat” variable, specifically (1 vs 0), is -8.172, but it is not statistically significant (z = -0.60, p = 0.550). This lack of significance suggests that, based on the available data, there is no robust indication of a substantial difference in violent crime rates between states with and without marijuana law liberalization.

# **Conclusion:**

This comprehensive analysis aimed to unravel the intricate relationship between marijuana law liberalization and violent crime rates across U.S. states. The examination involved a multifaceted approach, incorporating propensity score methods, logistic regression, exploratory data analysis through box plots, and outcome model estimation. The treatment selection model, guided by a carefully constructed Directed Acyclic Graph (DAG), considered crucial covariates such as historical marijuana policy stance, political ideology, religious affiliation, and the unemployment rate. Through propensity score matching, efforts were made to achieve balance between treated and control groups, essential for valid treatment effect estimation.

The exploratory analysis using box plots illuminated potential covariate imbalances, necessitating meticulous consideration in subsequent analyses. The logistic regression results revealed the nuanced influence of demographic and socio-economic factors on the likelihood of marijuana liberalization. The covariate overlap scatter plots provided visual evidence of common support between treatment and control groups, reinforcing the suitability of the matched samples for further analysis.

The outcome model, the DAG illustrated the complex causal relationships among covariates, highlighting the importance of controlling for baseline crime rates in 2014. The estimation of treatment effects using inverse probability weighted regression adjustment and simpler propensity score matching offered diverse perspectives. While the propensity score matching suggested a general reduction in violent crimes associated with marijuana liberalization, the inverse probability weight adjusted regression did not yield statistically significant evidence of a causal effect.

These findings underscore the complexity of the relationship between marijuana policy changes and violent crime rates, emphasizing the need for nuanced interpretations. The varied treatment effects across states and the lack of robust evidence in the inverse probability weight adjusted regression caution against sweeping generalizations. Contextual factors, regional dynamics, and unobserved variables may contribute to the diverse outcomes observed. This study contributes to the ongoing discourse on drug policy, providing empirical insights and paving the way for further nuanced investigations into the impact of marijuana law liberalization on societal well-being.

# **Appendix A: Do file for 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

\* 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 = real(violent\_crimepc\_2014)

gen crime\_2019 = real(violent\_crimespc\_2019)

\*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, over(treat)

graph box conservative, over(treat)

graph box catholic, over(treat)

\*Logistic regression for treatment selection model

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

unemployment\_2014 conservative catholic

estimates store logistic

\*Propensity scores

predict ps, p

\*roc

roctab treat ps

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

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

\*Histogram

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

fraction color(gs8) start(0) width(0.05)) ///

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

lcolor(black) start(0) width(.05)), ///

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

\* Covariate overlap scatter plot

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

(scatter unemployment\_2014 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 male\_2019 unemployment\_2019 young\_2019p crime\_2014) ///

(treat male\_2014 young\_2014p middle\_aged\_2014p unemployment\_2014 ///

catholic, logit)

\*Simpler propensity score matching

\*Inverse probabilit weights

gen ipw=1/ps if treat==1

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

\*Regression adjusted model for control and treatment groups

regress crime\_2019 male\_2019 unemployment\_2019 young\_2019p crime\_2014 ///

if treat==0 [aw=ipw]

predict pcontrol

regress crime\_2019 male\_2019 unemployment\_2019 young\_2019p crime\_2014 ///

if treat==1 [aw=ipw]

predict ptreat

\*Calculate ATU, ATT, ATE

gen teffect=ptreat-pcontrol

tabulate treat, summarize(teffect)

# **Appendix B: Log file for Analysis**

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

name: <unnamed>

log: C:\Users\mowni\Econometrics\_final\_project.log

log type: text

opened on: 28 Nov 2023, 21:43:25

. \*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)

(23 vars, 51 obs)

. \* 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 = real(violent\_crimepc\_2014)

(1 missing value generated)

. gen crime\_2019 = real(violent\_crimespc\_2019)

(1 missing value generated)

. \*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, over(treat)

. graph box conservative, over(treat)

. graph box catholic, over(treat)

.

. \*Logistic regression for treatment selection model

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

> unemployment\_2014 conservative catholic

note: conservative omitted because of collinearity.

Iteration 0: Log likelihood = -29.513972

Iteration 1: Log likelihood = -13.904908

Iteration 2: Log likelihood = -12.659163

Iteration 3: Log likelihood = -12.576693

Iteration 4: Log likelihood = -12.576509

Iteration 5: Log likelihood = -12.576509

Logistic regression Number of obs = 43

LR chi2(5) = 33.87

Prob > chi2 = 0.0000

Log likelihood = -12.576509 Pseudo R2 = 0.5739

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

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

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

male\_2014 | -109.463 89.22261 -1.23 0.220 -284.3361 65.41013

young\_2014p | -59.11765 37.94028 -1.56 0.119 -133.4792 15.24393

middle\_aged\_2014p | -145.2213 62.64728 -2.32 0.020 -268.0077 -22.43491

unemployment\_2014 | -.9122639 .5346818 -1.71 0.088 -1.960221 .1356933

conservative | 0 (omitted)

catholic | -32.82592 12.96631 -2.53 0.011 -58.23942 -7.412424

\_cons | 136.214 66.84857 2.04 0.042 5.193193 267.2348

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

. estimates store logistic

.

. \*Propensity scores

. predict ps, p

.

. \*roc

. roctab treat ps

ROC Asymptotic normal

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

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

43 0.9518 0.0319 0.88930 1.00000

.

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

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

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

treat

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

Male\_2014 -109.463

(89.223)

0.22

Young\_2014p -59.118

(37.940)

0.12

Middle\_aged\_2014p -145.221

(62.647)

0.02

Unemployment\_2014 -0.912

(0.535)

0.09

Catholic -32.826

(12.966)

0.01

Intercept 136.214

(66.849)

0.04

Pseudo R-squared 0.57

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

(collection ETable exported to file logisticmodel.docx)

. \*Histogram

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

> fraction color(gs8) start(0) width(0.05)) ///

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

> lcolor(black) start(0) width(.05)), ///

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

.

. \* Covariate overlap scatter plot

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

> (scatter unemployment\_2014 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 male\_2019 unemployment\_2019 young\_2019p crime\_2014) ///

> (treat male\_2014 young\_2014p middle\_aged\_2014p unemployment\_2014 ///

> catholic, logit)

Iteration 0: EE criterion = 1.370e-23

Iteration 1: EE criterion = 4.854e-26

Treatment-effects estimation Number of obs = 42

Estimator : IPW regression adjustment

Outcome model : linear

Treatment model: logit

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

| Robust

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

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

ATE |

Treat |

(1 vs 0) | -8.172256 13.66997 -0.60 0.550 -34.96491 18.6204

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

POmean |

Treat |

0 | 362.4172 23.12682 15.67 0.000 317.0895 407.745

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

.

. \*Simpler propensity score matching

.

. \*Inverse probabilit weights

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

(24 missing values generated)

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

(24 real changes made)

.

. \*Regression adjusted model for control and treatment groups

. regress crime\_2019 male\_2019 unemployment\_2019 young\_2019p crime\_2014 ///

> if treat==0 [aw=ipw]

(sum of wgt is 30.25911772251129)

Source | SS df MS Number of obs = 23

-------------------------+--------------------------------------------- F(4, 18) = 29.50

Model | 542007.364 4 135501.841 Prob > F = 0.0000

Residual | 82670.5374 18 4592.80763 R-squared = 0.8677

--------------------------+------------------------------------------- Adj R-squared = 0.8382

Total | 624677.901 22 28394.4501 Root MSE = 67.77

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

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

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

male\_2019 | 4610.12 1929.34 2.39 0.028 556.7275 8663.513

unemployment\_2019 | 32.53597 26.05236 1.25 0.228 -22.19801 87.26995

young\_2019p | 627.5722 1428.371 0.44 0.666 -2373.324 3628.469

crime\_2014 | .8928414 .1469413 6.08 0.000 .5841292 1.201554

\_cons | -2489.175 891.681 -2.79 0.012 -4362.528 -615.8233

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

. predict pcontrol

(option xb assumed; fitted values)

(1 missing value generated)

. regress crime\_2019 male\_2019 unemployment\_2019 young\_2019p crime\_2014 ///

> if treat==1 [aw=ipw]

(sum of wgt is 63.91014897823334)

Source | SS df MS Number of obs = 19

-------------------------+---------------------------------------- F(4, 14) = 54.31

Model | 97071.2319 4 24267.808 Prob > F = 0.0000

Residual | 6255.68283 14 446.834488 R-squared = 0.9395

-------------------------+---------------------------------- Adj R-squared = 0.9222

Total | 103326.915 18 5740.38415 Root MSE = 21.138

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

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

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

male\_2019 | 342.2181 1455.394 0.24 0.818 -2779.293 3463.729

unemployment\_2019 | -3.051153 15.45301 -0.20 0.846 -36.19457 30.09226

young\_2019p | 280.3521 913.3103 0.31 0.763 -1678.504 2239.208

crime\_2014 | 1.054977 .0981297 10.75 0.000 .8445099 1.265444

\_cons | -234.4558 778.5452 -0.30 0.768 -1904.269 1435.358

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

. predict ptreat

(option xb assumed; fitted values)

(1 missing value generated)

.

. \*Calculate ATU, ATT, ATE

. gen teffect=ptreat-pcontrol

(1 missing value generated)

. tabulate treat, summarize(teffect)

| Summary of teffect

Treat | Mean Std. dev. Freq.

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

0 | -14.257115 46.593689 23

1 | -1.2741932 45.18478 19

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

Total | -8.3838886 45.869679 42

.

end of do-file