Drinking to Crime Acts

Benjamin Cai

Econ 104

Professor Dobkin

Abstract

This paper analyzes the impact that the Minimum Legal Drinking Age (MLDA) has over crime rates in addition to reducing the proportion of the population that drinks alcohol. Two sets of data were collected through experiments done by the National Health Interview Sample Adult Files that dates from 1997 to 2007. The first data set contains information about the subjects ranging from their ethnicity, age, marital status to the number of days they drink alcohol in a week. The second data collected from arrest records set contains different types of crimes that individuals were arrested for and age. To answer the question of whether the MLDA reduces crime rates and alcohol consumption, we utilized regression analysis to perceive the effectiveness of the MLDA being set at 21. In addition, we created regression discontinuity graphs that are centered around age 21 as we are looking for the difference of crime rates and alcohol consumption before and after an individual turned 21 of age. Since the data of alcohol consumption and crime rates are from different data sets, we utilized a two instrumental variable approach to perceive if the MLDA influences lowering crime rates. Concluding from our analysis and graphs, setting the MLDA at 21 does reduce alcohol consumption below the legal drinking age while crime rates start to lower after the age of 21.

Introduction

There has been a debate whether the minimum legal drinking age is appropriately placed at 21 as some believe it should be lowered to 18 while others believe it's too young. Underage drinking has been a problem in society because young people tend to drink irresponsibly leading to violent acts or life-threatening situations. This tends to be the case especially in college campuses as there are a mixture of young adults that are of legal drinking age and those who are not that attend parties where alcohol is most likely served. To find a solution on whether the legal drinking age is appropriate, we used data set from a national survey done by the Centers for Disease

Control and Prevention that has information about an individual’s background information such as age, ethnicity, marital status, employment status along with percent of days they drink alcohol in a week. By creating regression tables and regression digression analysis graphs, we can see the difference of the population’s alcohol consumption and alcohol related crimes between after and before turning age 21. To conclude, is setting the minimum legal drinking age at 21 appropriate for reducing alcohol consumption and crime rates in the population?

With the first dataset, we can perceive that the survey encloses around teenagers and young adults as we want to center the regression discontinuity graph around age 21. By regressing these variables to their age, we can see if setting the MLDA to 21 would prevent more underage drinking and overall alcohol consumption throughout their 20s. On the other hand, we acquired another set of data that has information about an individual’s crime for arrest along with their age. List of crimes include assault, robbery, and alcohol related crimes that individuals were arrested for. We want to answer the question of whether alcohol consumption influences crime rates so we centered our digression on age 21 so then we can compare how the MLDA could affect crime rate before and after the legal alcohol consumption age. However, this can have omitted variable bias that can influence the results as crimes can be committed without the consumption of alcohol but from other causes were recorded into the dataset. In effect, we can utilize instrumental variable regression to cover some of the bias as we are looking for the direct cause of the arrest. By looking at alcohol related crimes, we can narrow our data to strictly look at the change in crime acts after turning 21.

After acquiring conclusive data from the regression analysis and data that were etched into scatter plots, we can assess the results drawn to whether the minimal legal drinking age influences reducing alcohol consumption and crime rates in the population. From the regression digression graphs, we can see that alcohol consumption rises and peaks at the age of 21 but declines steadily after the peak. In addition, for the crime rates, we can also see a steady decline in arrests due to all crime causes. For the regression tables generated, we can see that alcohol consumption does play a role in crimes committed as the coefficients are statistically significant for most variables. By regressing all the covariates against age, we can see that some covariates have a huge impact by the huge increase or decrease in the coefficient for age. With the analysis in mind, we can safely conclude that setting the MLDA at age 21 creates a difference in alcohol consumption and crime rate.

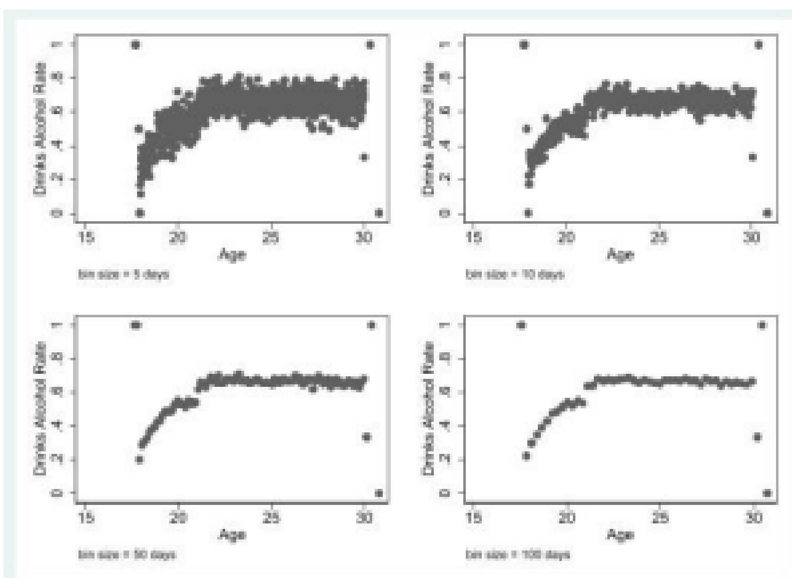
Data

The alcohol consumption dataset was collected from a National Health Interview Survey done by the Center for Disease Control and Prevention that was recorded in 1997 to 2007. This research is a form of survey where participants can choose to join the research and they answer questions regarding their alcohol consumption along with other simple categorical questions. Randomization was applied for this research as the CDC collected data from the population in a national level. Information regarding the individual includes their level of education, ethnicity, employment status, marital status, age, and if they drink alcohol. Even with all these variables, we are focusing on whether the subject has drunk alcohol before in their life. The reason for the vast number of categories in the data collected was to group the individuals into groups and form any kind of statistically significant conclusions that might arise from analyzing the data. For example, we might be able to conclude that some ethnicity in the population is more likely to drink alcohol than others. However, this research is susceptible to bias such as desirability bias and recall bias which leads to measurement error. For example, there are measurement errors in the variable “age” because young adults are more likely to consume alcohol due to easy access to parties and social gatherings. There can also be desirability bias because some subjects might lie, especially underage alcohol drinkers on whether they drink alcohol or not therefore skewing the population of people that consume alcohol to be lower. In addition, there can be recall bias on the percentage of days that a subject consumes alcohol because it is a long span of time to remember for some individuals. On the other hand, the arrest rate data set was collected from the Monthly Arrest and Citation Register from the state of California where it states the individuals arrested along with the category of their crime. The variable “all” contains the entirety of all arrests per 10,000 by age overall ranging from the different types of crime committed and arrested for. The type of violent crimes committed were labeled and put into subcategories that include drunk risk to self while drunk, driving under the influence, violation of liquor law, disorderly conduct and vagrancy, robbery, simple assault, and aggravated assault. With the usage of census data, researcher Dobkin and Carpenter was able to find the similarity and effect of alcohol consumption affecting crime rates. By gathering these variables, we can create regression tables and scatter plots to help us understand whether the MLDA is preventing underage drinking and overall crime rates in the nation.

Methods

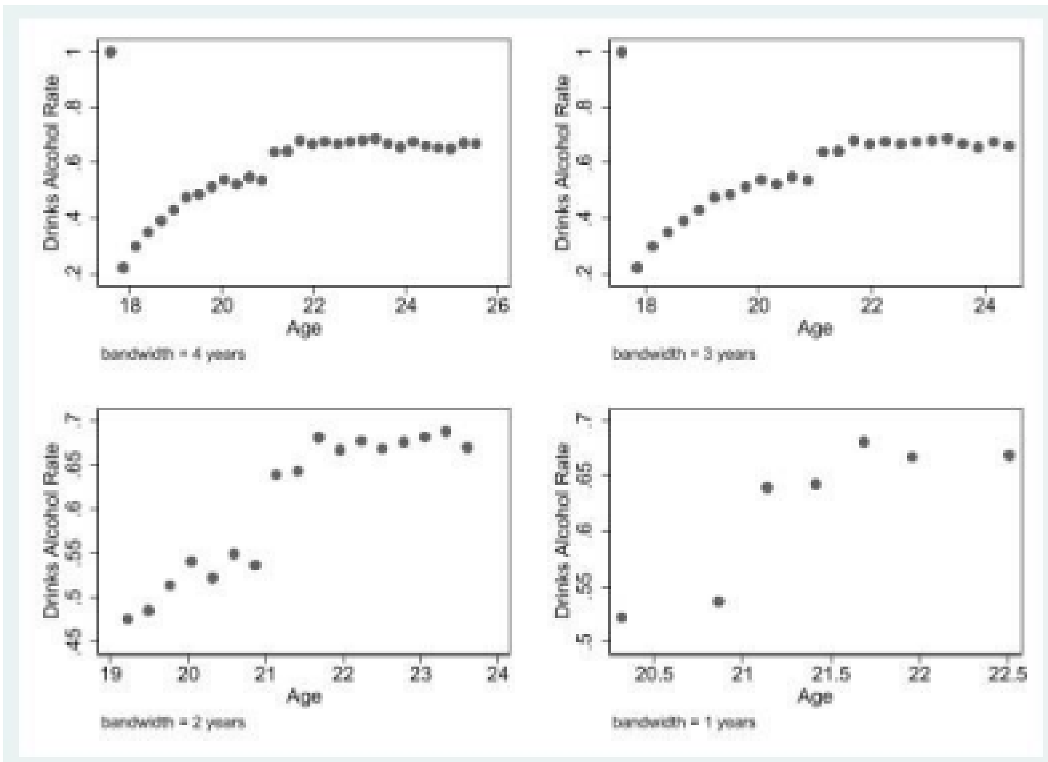
To analyze the discontinuity and see the difference of alcohol consumption at the minimal legal drinking age, we must create a clear graph that can be read and analyzed, we would need to design the regression discontinuity at the age of 21 for alcohol consumption and crime rate. With creating a good fit for regression discontinuity, we can clearly see the difference between the minimum legal drinking age and underage drinking. To do so, we must create a similar range of measurements for both variables allowing us to make an easier comparison. So, the graph is intended to show the age disparity and the percentage of days the individuals drink in a week. Therefore, I put their age on the x-axis and percent of days consuming alcohol on the y axis to create a scatter plot. In addition, to assemble a comprehensive scatter plot graph, we would have to adjust and pick the bin width and band width of the graphs. Bin width controls how condensed one section of the scatter plot can be so we would want to choose a bin width that minimizes the dispersion of the data and creates a less noisy scatter plot that is readable.

Figure 1. Bin Width Selection



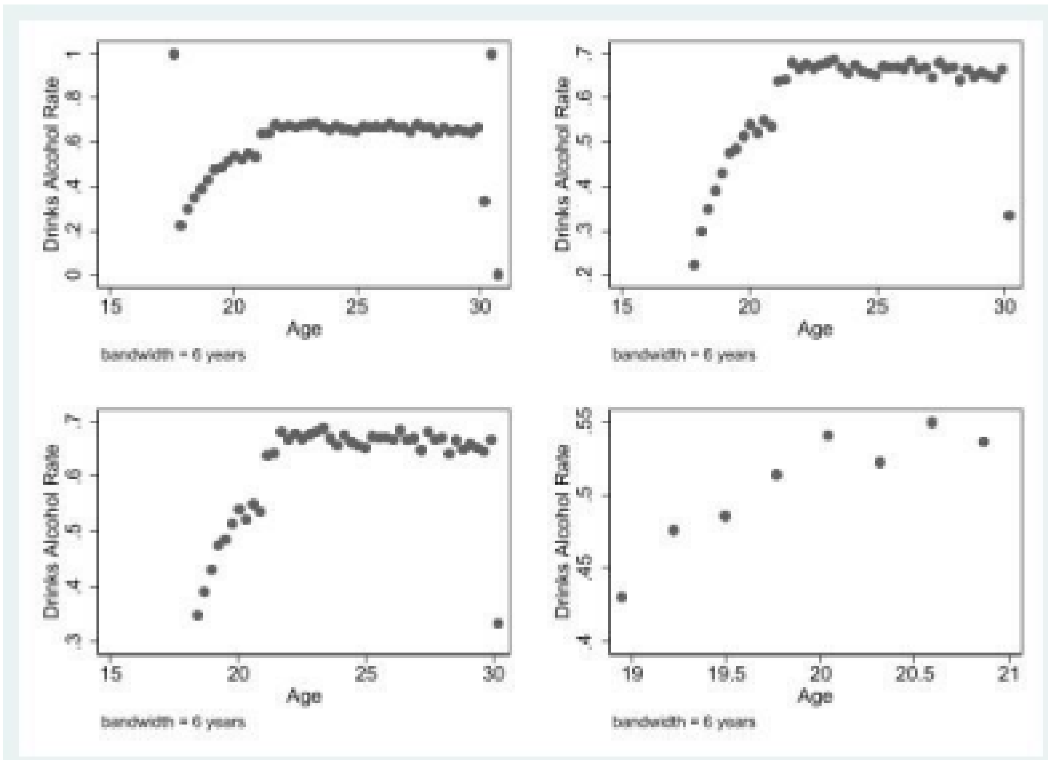
As Figure 1 demonstrates, we can see that I have set the bin size for 5, 10, 50, 100 days. By increasing the bin size, we are conglomerating more data into one point on the scatter plot. In general, we are combining some of the similar individuals that are the same age and consume relatively the same amount of alcohol consumption into one point on the regression discontinuity scatter plot. So, in this scenario, I picked the bin size of 50 days because it minimizes the dispersion while maintaining a regression linearity. In addition, the 50 days bin size makes the scatter plot less messy compared to a bin size of 5 or 10 days. Otherwise, as shown on bin size 5 and 10 days, data are too cluttered in which we cannot analyze nor find a good line of regression. In addition, the bin size of 100 days would show far too less data for us to depict a regression linearity. Next, I chose the band width which is the x plot of the graph in which by decreasing the bandwidth, we would be decreasing the plots of the graph however summing up the same age individuals into a single point on the plot.

Figure 2. Bandwidth Selection



As shown on Figure 2, we can see that if we decrease the bandwidth in terms of years, there seems to be less points on the scatter plot meaning we are removing some measurement of ages on the x plot. As shown, bandwidth of 4 years includes ages 18 to 26 skipping the odd numbers while bandwidth of 1 year is included every year in addition to half years. In this case, I picked the bandwidth of 2 years because it sorts the age groups to be more general while retaining the regression linearity. After picking my bin width and bandwidth, I adjusted the range of my y-axis to exclude any outliers because those are extreme cases and will affect the result of our conclusion.

Figure 3. Y-Range Selection



Shown on Figure 3, I expanded and decreased the y range for the scatter plot which gives us a different depiction on the scatterplot. In this dataset, I picked my y-range to be from .2 to .7 because it alienates the outliers while maintaining the rest of the data to give us an accurate representation of data plotted.

In addition to the graphs, I also implemented a two stage least squares regression analysis which contains the first stage (FS) and reduced form (RF) to combine the cause and effect of drinking alcohol directly affecting the crime rates. To summarize, turning 21 will allow the individuals to purchase and consume alcohol which in effect can increase the arrest rate due to alcohol being the underlying cause. To calculate if that statement is true, we would use this equation:

$$Y_i = \beta_0 + \beta_1 Over\ 21_i + g\ (Age_i\ ) + \varepsilon_i$$

By using this regression equation, we can regress their age using any line of best fit such as linear, quadratic, cubic to sum up the data while giving us an exact number to generate the exact difference between the legal and nonlegal consumption of alcohol. By putting the coefficient “Over21” first in the regression equation, we can see if other variables are affecting the effect of being over the age of 21 with other coefficients that were omitted before. With the variable “Over21”, we are looking for the effect of turning 21 and being exposed to new policies that can change how the individuals react such as being able to purchase alcohol. In addition, I created First Stage and Reduced Form equations to fit the regression line of best fit better for each of the sub causes and alcohol consumption.

$$FS: Alcohol_i = \rho_0 + \rho_1 Over\ 21_i + g\ ((Age - 21)_i\ ) + \leq_i$$

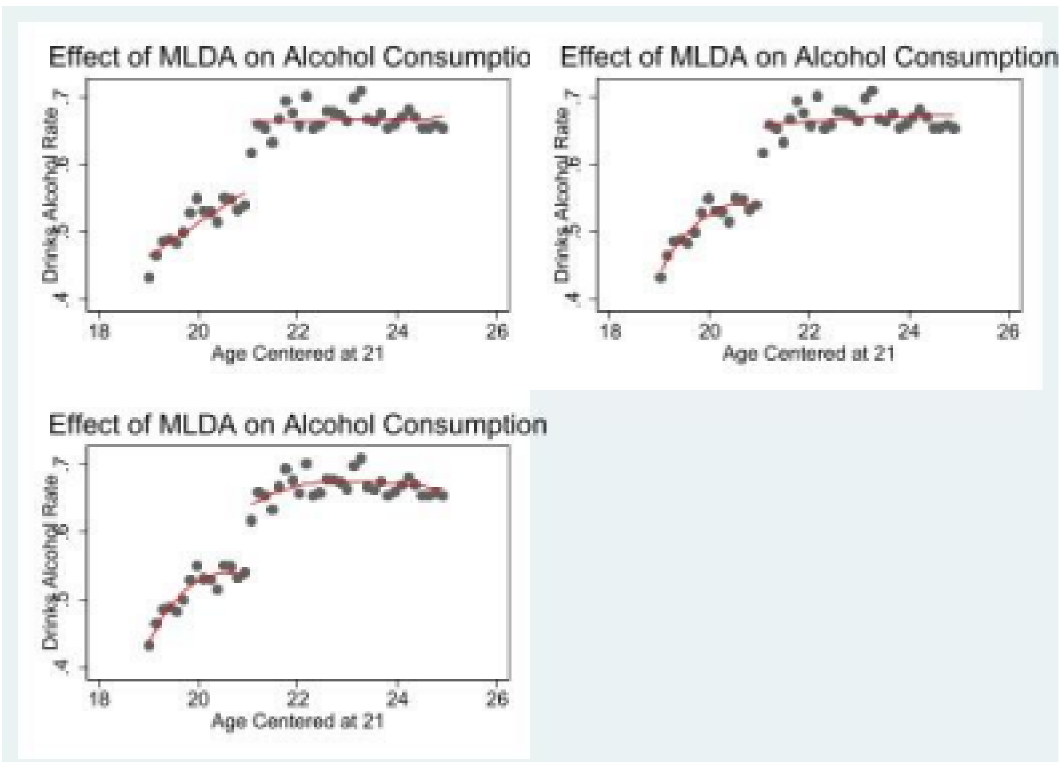
$$RF: Arrest_A = \varphi_0 + \varphi_1 Over\ 21_A + h((Age - 21)_A) + \eta_A$$

Through the First Stage equation, I can regress the individuals being over and under the age of 21 and how it affects the alcohol consumption. With the Reduced Form equation, I regressed the effect of being over and under the age of 21 and how it affects the individuals getting arrested for crimes committed. By combining these two equations and utilizing the delta method, I am trying to bridge the connection of how likely it is for someone turning 21 and consuming alcohol and in effect having them committing crimes. Through these methods, I generated regression discontinuity scatter plots and regression tables that helped me create a statistically significant conclusion.

Results

From generating and adjusting the ranges for our graphs, I created the results for the regression discontinuity graphs for alcohol consumption along with the individual's age.

Figure 4. Regression Lines



As Figure 4 is showing, I incorporated three different types of line of best fit containing linear, quadratic, cubic regression. Based on the three choices, I chose the cubic regression line of best fit because it sums up the data for both parts of the data sufficiently while giving us the difference between legal and nonlegal alcohol consumption. To reinforce the idea being perceived, I have created regression tables that include the age in different regressions against the drinking percentage per week variable.

Regression Table 1

(1) (2) (3) (4) (5) (6)

Drinks Alcohol Drinks Alcohol Drinks Alcohol
VARIABLES Drinks Alcohol Drinks Alcohol Drinks Alcohol

Over 21 0.0861*** 0.0861*** 0.0916*** 0.0910*** 0.0805*** 0.0784*** (0.0141) (0.0142) (0.0213) (0.0215) (0.0287) (0.0291) Age - 21 0.0439*** 0.0439*** -0.0236 -0.0236 -0.0509 -0.0509 (0.00896) (0.00897) (0.0358) (0.0358) (0.0899) (0.0899) (Age – 21) *Over
21 -0.0237* -0.0236* 0.0937* 0.0949* 0.214* 0.222* (0.0123) (0.0123) (0.0493) (0.0496) (0.124) (0.125)
(Age-21) ^2 -0.0340* -0.0340* -0.0681 -0.0681 (0.0175) (0.0175) (0.104) (0.104)
(Age-21) ^2 *
Over21 0.00924 0.00873 -0.0723 -0.0801 (0.0239) (0.0240) (0.143) (0.145)
(Age – 21) ^3 -0.0114 -0.0114 (0.0343) (0.0343)
(Age – 21) ^3 *
Over 21 0.0498 0.0520 (0.0471) (0.0474)
Birthday 0.00238 0.0198 0.0370 (0.0838) (0.0845) (0.0856)
Constant 0.559*** 0.559*** 0.536*** 0.536*** 0.532*** 0.532*** (0.0102) (0.0102) (0.0153) (0.0153) (0.0207) (0.0207)

Observations 18,824 18,824 18,824 18,824 18,824 18,824 R-squared 0.025 0.025 0.025 0.025 0.025 0.025
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To conclude from Figure 4 and Regression Table 1, the drinking rate is 12.5% lower for people directly under the MLDA than it is for people above it. In addition, the regression table tells us that by adding the cubic regression cancels out most of the bias therefore concluding that the cubic regression is the best fit to summarize the graph and data in general.

To get rid of more of the omitted variable bias, I included the demographic variables of the individuals such as their ethnicity, marital status, and others to potentially create conclusions that some demographics tend to drink more alcohol.

Balance Table 2

(1) (2) (3) (4) (5) (6)
VARIABLES HS Diploma Hispanic white black married employed

Over21 0.011 -0.010 0.016 -0.014 -0.030*** 0.008 (0.011) (0.012) (0.014) (0.010) (0.010) (0.014)
Constant 0.821*** 0.241*** 0.554*** 0.157*** 0.152*** 0.642*** (0.008) (0.009) (0.010) (0.007) (0.007) (0.010)

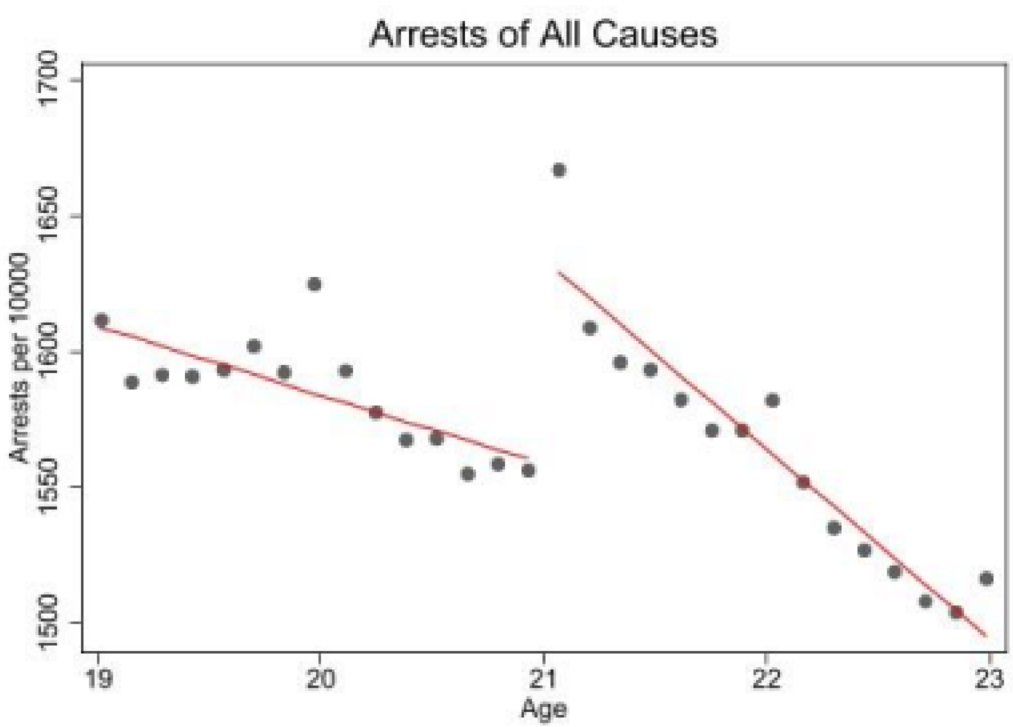
Observations 18,824 18,824 18,824 18,824 18,824 18,824 R-squared 0.003 0.000 0.000 0.000 0.018 0.014 Standard errors in parentheses
*** p<0.01, ** p<0.05, *

p<0.1

According to Balance Table 2, we can see for all the characteristic variables, are all continuing as before the age of 21 and after. This shows that our data are covering all before and after the MLDA. In effect, we can conclude that these variables are not specific to over or under the MLDA but are consistent throughout the age of 19 to 23.

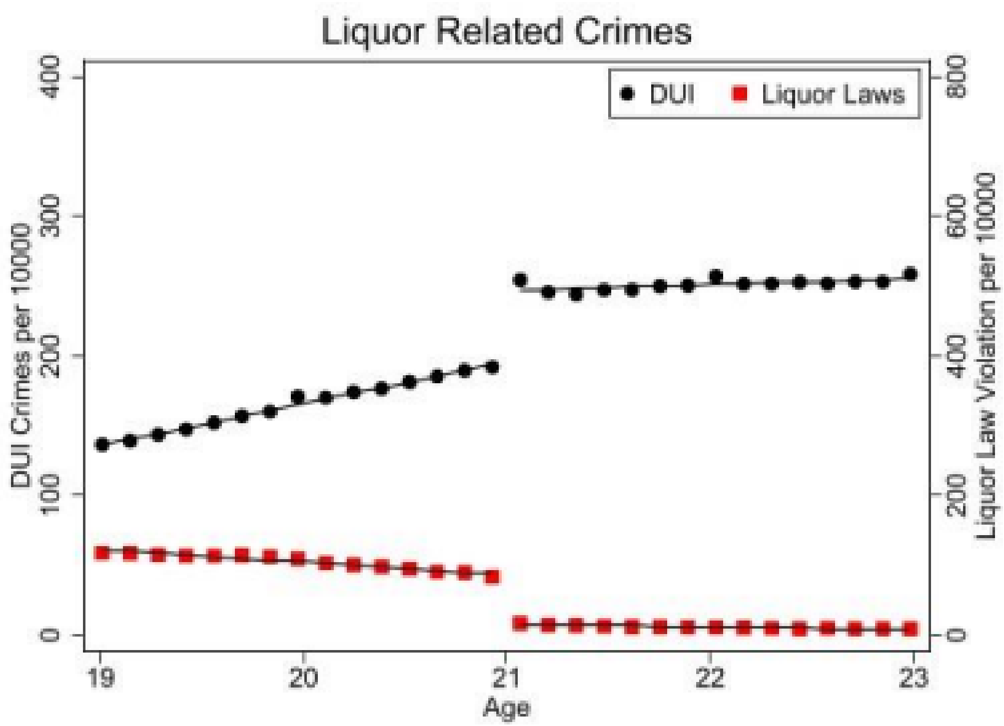
For our crime data, I created multiple regression discontinuity scatter plots because there are many forms of crime that the individuals were arrested for. However, there is a variable that sums up all the arrests in which I made a R-D scatter plot to show the difference before and after the MLDA.

Figure 5



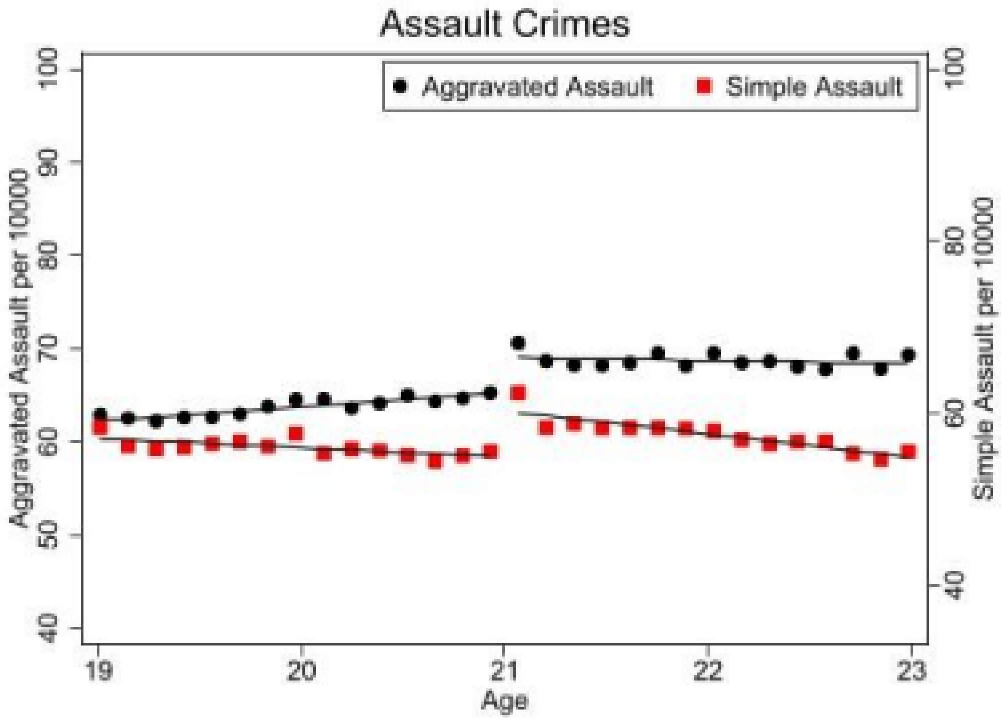
Shown on Figure 5, we can see that arrests were fluctuating between 19 and 21 but starts go decline towards age 21. However, the arrest rate spikes dramatically at 21 but has a sharper decline towards age 23. This could potentially mean that as individuals get older, they tend to avoid any criminal activities that could get them arrested. The difference between the under the MLDA and over is around 1,000 per 10,000. Since that number is huge, we would want to separate and categorize the types of crimes in which we can eliminate some bias that some crimes committed are unrelated to alcohol consumption. To categorize the types of crimes, I created additional arrest rates for the specific crime committed ranging from alcohol law violations to aggravated assault.

Figure 6



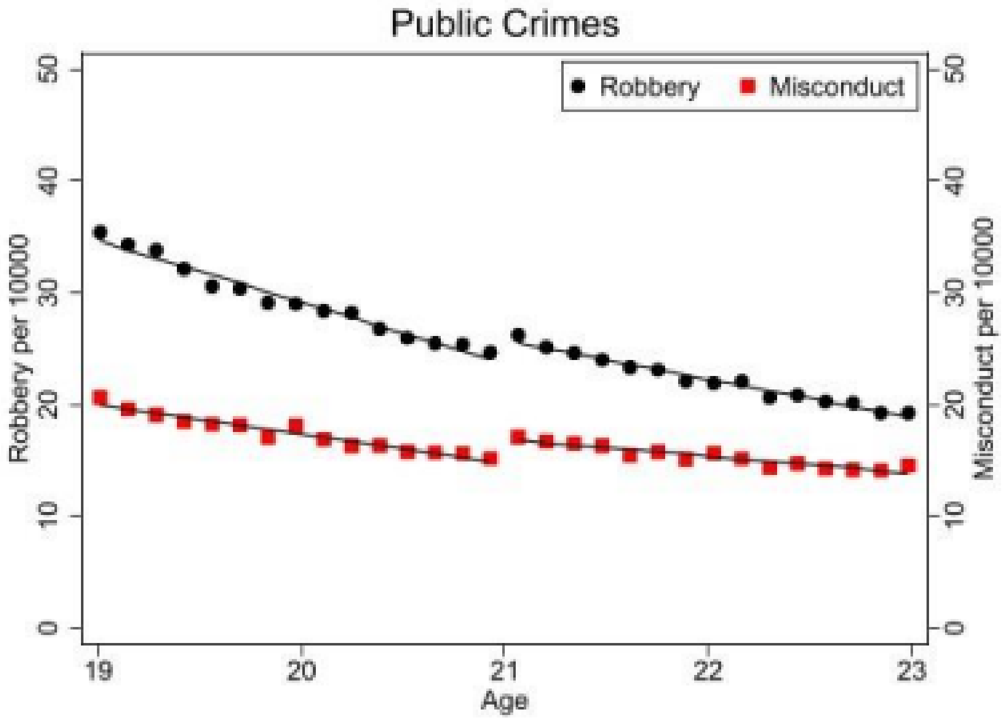
As shown on this R-D scatter plot, driving under influence crimes were at a steady incline and stayed around the same level after MLDA towards age 23. On the other hand, violation of liquor laws dropped drastically after the MLDA largely because liquor laws usually center around preventing underage drinking. To explain further, people above the MLDA will not be able to violate most of the liquor laws. Next, I implemented the two different kinds of assault crimes in which individuals were arrested for another R-D scatter plot.

Figure 7



For aggravated assault and simple assault, we can see that they both increase after the MLDA; however, the correlation between alcohol consumption and assault crimes are unknown. This is because assault can simply be an act of violence against somebody or property without the influence of alcohol. Lastly, I created an R-D scatter plot for societal crimes such as robberies and misconducts or any kind.

Figure 8



Shown on the Figure 8 scatter plot, both robberies and misconducts are at its peak at the age of 19 but slowly decline towards age 21 and 23. This can be explained that small levels of robberies usually do record the robber because they are usually not caught. So, this can be measurement error in some sense as uncaught robbers were not included in the dataset. In addition, misconduct is a broad category that has multiple layers of offense, alcohol consumption might have a minuscule effect on the outcome of our misconduct data. To test the validity of these data, I created a regression table for all the crime categories to ensure that we can create a statistically significant conclusion based on our scatter plots.

(1) (2) (3) (4) (5) (6) (7) VARIABLES All Crimes All Crimes DUI DUI Liquor laws Liquor laws Robbery

Over 21 -28.86*** -33.90*** 41.60*** 40.97*** -96.62*** -96.92*** -1.971*** (4.478) (4.556) (0.947) (0.966) (0.816) (0.834) (0.239) Age - 21 88.12*** 88.12*** 46.04*** 46.04*** 5.574*** 5.574*** -2.594*** (1.370) (1.364) (0.290) (0.289) (0.250) (0.250) (0.0731) (Age-21) * Over21 -161.2*** -159.4*** -48.62*** -48.38*** -7.630*** -7.520*** 0.0824 (1.938) (1.961) (0.410) (0.416) (0.353) (0.359) (0.103) Birthday 60.37*** 7.644*** 3.535* (11.29) (2.395) (2.067) Constant 1,665*** 1,665*** 211.9*** 211.9*** 110.3*** 110.3*** 26.80*** (3.168) (3.153) (0.670) (0.669) (0.577) (0.577) (0.169)

Observations 2,922 2,922 2,922 2,922 2,922 2,922 2,922 R-squared 0.705 0.708 0.971 0.972 0.944 0.944 0.815

(8) (9) (10) (11) (12) (13) (14) robbery Aggravated assault Aggravated assault Sim assault Sim assault Drunk risk Drunk risk

-2.145*** 1.761*** 1.537*** 11.64*** 11.31*** 16.46*** 14.50*** (0.244) (0.338) (0.345) (0.462) (0.471) (0.900) (0.903) -2.594*** 3.954*** 3.954*** -6.525*** -6.525*** 17.04*** 17.04*** (0.0729) (0.103) (0.103) (0.141) (0.141) (0.275) (0.270)
0.147 -4.628*** -4.545*** 3.925*** 4.045*** -22.55*** -21.82*** (0.105) (0.146) (0.148) (0.200) (0.203) (0.390) (0.389) 2.090*** 2.686*** 3.886*** 23.48*** (0.604) (0.854) (1.168) (2.239) 26.80*** 67.63*** 67.63*** 48.51*** 48.51*** 118.0*** 118.0*** (0.169) (0.239) (0.239) (0.327) (0.326) (0.637) (0.625)

2,922 2,922 2,922 2,922 2,922 2,922 2,922 0.816 0.575 0.576 0.530 0.532 0.804 0.811

(15) (16)
Misconduct Misconduct

5.051*** 4.976***
(0.216) (0.221)
-5.240*** -5.240***
(0.0662) (0.0662)
3.763*** 3.791***
(0.0936) (0.0952)
0.898
(0.548)
11.75*** 11.75***
(0.153) (0.153)

2,922 2,922
0.814 0.814

In this regression table, I included all the categories of crime that were graphed in my regression table. In this regression table, we can see they are statistically relevant at the 99% confidence level because the asterisk behind each coefficient signifies that each covariate's p-value is under .01. To summarize, all these covariates are statistically significant in relation to being over the MLDA and committing these crimes. However, the huge increases during the age of 21 can be partially due to celebrating their birthday and in effect can potentially get them in trouble with the law. From these regression discontinuity scatter plots and regression table analysis, we can assess whether the

MLDA is effective in lowering crime rates and underage alcohol consumption based on the difference of before and after the age of 21.

Conclusion

Through all our regression tables and regression discontinuity scatter plot examining the data of alcohol consumption based on survey response and concrete data about arrest rates, we can utilize them to answer the question of whether the minimum legal drinking age is set appropriately at age 21 to prevent underage drinking and lower crime rates. For the research to be implemented correctly, there are three assumptions that need to be fulfilled.

Assumption 1: The assumption that the sample is randomly assigned and unbiased is true because the CDC compared the same population of individuals but at separate time periods.

Assumption 2: When an individual turns 21, they usually experiment with alcohol since it is legal for them to purchase at that point. In effect, this assumption holds as turning 21 affects whether they drink alcohol or not.

Assumption 3: Through our analysis, we cannot say that consuming alcohol is the sole reason people are committing more crimes when turning 21. So this assumption cannot hold as other factors are causing higher crime rates.

To connect the relevance between alcohol consumption and crime rates, I utilized instrumental variables that are placed into one dataset so I can IV regress the two coefficients to test for statistical significance. From the IV regression, I got a coefficient (695.778) and standard error (117.413) in which you divide each other to get the t stat. The t-stat came out to be greater than 99% confidence level so the effect of alcohol consumption will potentially affect certain categories of crime rates. So, my conclusion would be that it does not prevent but lower the percentage of underage alcohol consumption because there will always be teenagers that are curious or have access to alcohol so those parameters cannot be prevented. The MLDA is not powerful enough to stop every 18-year-old attending college to drink alcohol as it is easily accessible at parties and social gatherings. On the other hand, the MLDA has slim to no effect on lowering crime rates such as robberies, assault, and misconduct after the age of 21 because they do not require the consumption of alcohol for the individuals to commit these crimes. However, they have a slight impact on DUIs and Liquor law violations because

They are alcohol related crimes. To sum up, it does not matter what age the MLDA is set at, because there will always be an influx of alcohol consumption on the individual's 21st birthday and beyond. However, setting the MLDA at age 21 can lower the percentage of underage drinking which can prevent any growth development hindrance to the teenagers.