

The Effect of the Minimum Legal Drinking Age on Alcohol Consumption and Arrests for Young Adults

Thipbadee Tantivilaisin

12 June 2019

Abstract

Recently there has been a large legislative lobbying effort by many nonprofit organizations to lower the minimum legal drinking age (MLDA) in the United States. We quantify the effectiveness of the MLDA by measuring how well it prevents underage drinking and arrests. The datasets were retrieved from two separate sources, the first being the National Health Interview Survey (NHIS) for alcohol consumption and California's Monthly Arrest and Citation Register (MACR) for arrest information. Using a regression discontinuity design, we find that the MLDA reduces the amount of drinkers by 9.1 percentage points and arrests rates by 82.73 per 10,000 person-years. These findings illustrate that the MLDA has a significant impact on the reduction of drinking and crime among young adults.

1 Introduction

The minimum legal drinking age (MLDA) is a federal law that specifies the age threshold that an individual must be in order to legally purchase alcohol. This is currently set at 21 years in the United States. This law was intended to reduce the harms resulting from alcohol consumption such as increased internal health risks, arrests, and ultimately death. However, in recent times there are nonprofit organizations such as the Amethyst Initiative that are putting in a considerable amount of effort in trying to reform the MLDA to a younger age. The group postulates that the current MLDA creates a dangerous and clandestine culture of drinking on college campuses that would otherwise not exist if the MLDA were lower. So does having the MLDA set at age 21 reduce the number of young adults consuming alcohol and does it reduce the harms caused by it such as arrests? This question is motivated by the fact that changing the MLDA would substantially affect the way in which young adults are able to obtain alcohol, which may ultimately lead to a significant increase in harm (we will quantify this with arrests) to this target population.

Our alcohol consumption data is drawn from the National Health Interview Survey (NHIS) Adult Files from 1997 to 2007. The NHIS is a data collection program ran under the Centers for Disease Control designed to report the health of the United States population. The survey includes race, insurance status, employment status, age, drinking behaviors, and whether the individual has consumed alcohol in the past. Our crime data comes from California's Monthly Arrest and Citation Register (MACR) from 1979 to 2006. This dataset includes the type of arrest and how many days away the individual is from their 21st birthday.

We take advantage of the fact that as an individual turns 21 in the United States, the only change in the individual is their ability to legally purchase alcohol. This gives us two distinct groups of individuals,

one just below 21 who cannot purchase alcohol (control group) and individuals 21 or older who can (treatment group). Thus, we are able to use a regression discontinuity design to approximate the effect of being subjugated to the MLDA on arrests and alcohol consumption by comparing the two very similar groups. We use age as the running variable and 21 years of age as the threshold.

With this design we observe a statistically significant increase the percentage of individuals who consume alcohol in addition to the number of arrests. We approximate that the MLDA reduces the number of drinkers by 9.1 percentage points. Furthermore we find that the MLDA reduces the number of arrests by 82.73 per 10,000 person-years for all types of arrests. When looking at each category of arrests, we observe that those not subjugated to the MLDA are more likely to be arrested for every type of crime excluding liquor laws. These results suggest that the MLDA does reduce the number of young adults that consume alcohol and harm being done to the target population.

2 Data

The data that we use for alcohol consumption is drawn from the National Health Interview Survey (NHIS) Adult files ranging from 1997 to 2007. The NHIS is an annual survey designed by the Centers for Disease Control in 1957 to represent the overall health of non-institutional civilian Americans (citizen and non-citizen) and it is administered by the U.S. Census Bureau. The sampling is divided into four panels and hour long interviews with approximately 35,000 households and 87,500 individuals. Children and adults not home during the interview may have their information provided by another family member who is at least 18 years of age or older. Specifically for the Adult files only one adult per household is selected at random and their basic information on health status, health care services, and health behaviors are collected. For our purposes, we are using information from the Adult files on drinking behavior and basic demographic characteristics. We use the binary variable Drinks Alcohol to signify whether the individual does so or not.

Some concerns that must be accounted for in this dataset is recall bias, desirability bias, oversampling, and measurement error. With regards to recall bias, it should not be a large concern other than an individual's drinking behavior such as drinking frequency. Desirability bias, however, attenuates our estimates of how many individuals under 21 drink, as drinking is illegal and therefore considered socially undesirable behavior. By 1997 the NHIS began oversampling black and Hispanic households to increase the precision of estimates for these two groups and by 2006 this also expanded to the Asian population. Thus, individuals from these groups were weighted much more in the survey as opposed to their true proportion of the U.S. population. Lastly, due to the way that we calculated an individual's date of birth, some individuals who are 21 years old are misrepresented as 20 and individuals who are 20 are represented as 21. This systematically reduces our estimates because the group to the left of the threshold will be brought up by 21 year olds who are misrepresented. The group to the right of the threshold will also be brought down by individuals are are less than 21. Overall, this will decrease the size of the jump at the threshold.

The data we use for arrests comes from California's Monthly Arrest and Citation Register (MACR)

from 1979 to 2006. The MACR collects felony and misdemeanor arrests for individuals in addition to race/ethnicity, sex, date of arrest, and age down to the date of the arrest. For our purposes, we will only be utilizing type of arrest and age of the individual at the time of the arrest. We also calculate the arrest rates per 10,000 person-years by dividing the total number of arrests at each age by the total number of individuals at that age (derived from the census). There are not too many concerns for this dataset as it is from administrative records, however we must be cognizant of that data entry errors may occur as well.

3 Methods

Our regression discontinuity design takes advantage of the fact that as an individual turns 21 years old, there is a discrete difference in their ability to legally purchase alcohol. Furthermore, this is assumed to be the only significant change that an individual undergoes as a direct result of turning 21 years old in the United States. Thus, this gives us two very similar groups that we can use to observe the effect of the MLDA on alcohol consumption and arrests. These two groups are individuals just under 21 years of age that is subject the MLDA, and individuals 21 or older who are not. We use a graphical approach for our regression discontinuity design to display the trends of our outcome variables of Drinking Rate and Number of Arrests per 10,000 person-years with Age as our running variable.

By running two lines of best fit on both sides of our threshold, we can easily observe any significant jump that occurs at age 21. For the bin-width to display in our figures, we chose to use a 30 day bin-width. When deciding this, we had to consider that we wanted to keep as many data points as possible to truly represent the data but we also wanted to turn down the noise enough so we can see the trend happening in our design. In Figure 1 we observe six different bin-widths for the drinking profile. The 30-day bin-width was chosen due to how much noise it reduced while preserving as many data points as possible.

For the bandwidth, we chose from age 19 to 23. In Figure 2 we can observe the various bandwidths for the drinking profile. With regression discontinuity, the most important aspect of the figure is at the discontinuation. However, one must still be able to see the trend in the outcome variable. The bandwidth of age 19 to 23 was chosen due to how it displayed the trend of the outcome variable while still displaying a clear jump at the threshold.

In Figure 3, we observe the various Y-Ranges for Drinking Rate. When choosing the proper range we want to have it such that the data points are centered within the graph in addition to the range being large enough that the overall trend in the outcome variable is visible. We chose the range of 45 to 70 percentage points because it centers the data points in our graph and it is large enough to illustrate a discrete jump at the threshold.

In Figure 4, we observe the various polynomial orders for the lines of best fit in our regression discontinuity design. When choosing the correct order for the polynomial, we want the model to be able to capture the curvature of the trend without overfitting the data. We found the quadratic fit to be the best balance between these requirements.

The regression model that we run to estimate the effect of the MLDA for both the Drinking Rate and Arrest Rate per 10,000 person-years is the following:

$$Y_i = \beta_0 + \beta_1 MLDA_i + f(Age_i) + \epsilon_i \quad (1)$$

where Y_i is age specific for Drinking Rate or Arrests per 10,000 person-years for all types of arrests. $MLDA$ is a binary variable such that if an individual is 21 years old or older it is equal to 1 and 0 otherwise. The coefficient of interest that captures the effect of the MLDA at the threshold is embodied by β_1 . $f(Age_i)$ is a quadratic polynomial such that Age_i interacts with the $MLDA$ binary variable. Furthermore, we also add *Birthday* when the outcome is Arrests per 10,000 person-years to account for the celebration effect. This effect causes an outlier within our dataset, such that individuals are much more likely to be arrested on their birthday than any other day. Thus, we add this binary variable to drop these outliers to illustrate that the effect of the MLDA we are seeing is not due to the celebration effect.

Another tool that we utilize is the Delta Method. This can be used to approximate the effect of the MLDA on Arrests per 10,000 person-years for individuals who abstained from drinking prior to turning 21 and partake in drinking after turning 21. The equations are as follows:

$$DrinksAlcohol_i = \phi_0 + \phi_1 MLDA_i + f(Age_i) + u_i \quad (2)$$

$$Arrests_A = \rho_0 + \rho_1 MLDA_A + f(Age_A) + \eta_A \quad (3)$$

where $DrinksAlcohol_i$ is the rate on whether an individual drinks alcohol. $MLDA$ is once again a binary variable such that if an individual is 21 years old or older and it takes on the value of 1 and 0 if not. ϕ_1 captures the effect of the MLDA on the drinking rate for the individual. $f(Age_i)$ is a quadratic polynomial such that Age interacts with the $MLDA$ binary variable.

$Arrests_A$ is Arrests per 10,000 person-years for all types of arrests and can be substituted for other subcategories of arrests as well. $MLDA$ takes on the same role as Equation 2 but for each age. ρ_1 is the effect of the MLDA on Arrests per 10,000 person-years at each age. $f(Age_A)$ is a quadratic polynomial such that Age interacts with the $MLDA$ binary variable.

We use Equations 2 and 3 to derive the instrumental variable coefficient:

$$\tau_{iv} = \frac{\rho_1}{\phi_1} \quad (4)$$

where τ_{iv} is the effect of the MLDA on Arrests per 10,000 person-years for individuals who abstain from drinking prior to turning 21 and partake in drinking after turning 21.

4 Results

Before we can draw any conclusions from our regression discontinuation model, we must verify that individuals under 21 years old are similar to individuals 21 and over. Thus, we can assume that the only difference between these two groups is that individuals above 21 are able to legally purchase alcohol.

If this assumption holds true, then we are able to isolate the effect of the MLDA on arrests and the proportion of individuals that drink. In Table 1, we observe the characteristics of the individuals which are the following: insurance status, if they have a high school level education, race, marital status, whether they attend school, race, and employment status. The difference between the two groups is embodied in the coefficient *MLDA* for each of the columns. In order to test that all of these differences are statistically insignificant, we assume the worst case scenario of sinister dependence. Thus, we divide our significance level by the number of tests, and get $\frac{0.05}{10} = 0.005$. We compare 0.005 to the p-values of the MLDA coefficient for all 10 tests and find that none of p-values are less than 0.005.

Now that we have verified that the differences between the two groups are statistically insignificant, we can now isolate the effect of the MLDA on the proportion of individuals that drink and the Arrest per 10,000 person-years. Figure 5 provides strong evidence of a discontinuous relationship between age and the percentage of the population that drinks alcohol. This discontinuous relationship can be observed at the threshold age of 21 which is exactly when the MLDA no longer applies. Table 2 captures the numerical relationship between age and the drinking rate that is graphically presented in Figure 5. In Table 2, we can observe the effect of the MLDA on the percentage of individuals that drink alcohol through the MLDA coefficient which accounts for the potential increase or decrease in the drinking rate in our model after an individual turns 21 years old. We observe from Table 2 that the MLDA coefficient gives us a statistically significant increase in the proportion of individuals that drinks alcohol by 9.1 percentage points when not accounting for the celebration effect and 9.2 percentage points when we are not. Thus, we can extrapolate that the MLDA is reducing the number of young adults that are drinking alcohol due to the graphical discontinuity of the relationship between age and drinking rate when age is 21 illustrated in Figure 5 and the statistically significant MLDA coefficient in Table 2.

In Figure 6 we observe a clear discontinuation in the relationship between age and Arrests per 10,000 person-years when age is 21 which is when individuals are no longer subject to the MLDA. This discontinuation is embodied by the MLDA coefficient in column one of Table 3, and it illustrates that the MLDA reduces the number of Arrests per 10,000 person-years by a statistically significant 82.73 percent. This trend continues in Figure 7 when we observe the relationship of age and the subcategories of Arrests per 10,000 person-years. The MLDA has a statistically significant effect in increasing the number of arrests for all subcategories except for liquor laws which is presented in columns two to eight in Table 3. This exception can be explained by the fact that most arrests as a result of violating liquor laws is due to underage drinking, which obviously cannot happen if one is legally allowed to consume alcohol. In Table 4 when we control for the celebration effect we reach the same conclusions as Table 3 for all subcategories. This evidence suggests to us that the MLDA is working in terms of reducing the number of young adults that are being arrested.

5 Conclusion

By verifying that the individuals around the threshold were similar in terms of their observable characteristics, we were able to conclude that the only impactful difference between the two groups are that

one can legally purchase alcohol and the other cannot. Thus, we are able to isolate the effect MLDA on the proportion of individuals that drink and Arrests per 10,000 person-years. Visually we found that at the threshold age of 21, there is a large discontinuity between the relationship of age and the respective outcomes. When these discontinuities were represented numerically, we found that immediately after individuals are no longer subject to the MLDA there is a statistically significant difference between individuals just over and under 21 in terms of the respective outcomes. This provides us with substantial evidence that the MLDA is indeed working as intended as it is reducing the number of young adults that are consuming alcohol and being arrested.

With the delta method which was mentioned in section 3, we can derive the local average treatment effect (τ_{iv}) of drinking alcohol on arrests for the set of individuals that abstain from drinking before turning 21 and partake in drinking after they turn 21. The results are summarized in Table 5 in the IV Estimate row. We observe that these estimates are all statistically significant and takes on the value of 866.46 arrests per 10,000 person-years for all arrests or a 59.31 percent increase. However, for us to accept these estimates the following three assumptions must hold. The first stage must be greater than zero, the MLDA only affects the number of arrests through an individual's alcohol consumption, and the assignment to treatment must be just as good as random.

To verify that the first stage is greater than zero in our model, we must know the number of individuals who began drinking as a result of being legally allowed to. This is trivial, as we have already verified this in section 4. Visually, the first stage is the distance of the discontinuation in Figure 5. This distance is captured in the MLDA coefficient in table 2 and it comes out to 9.1 percentage points.

The next assumption that must hold is that the MLDA only affects arrests through alcohol consumption. This assumption fails as individuals are not only allowed to drink alcohol, but they are also able be in places that serve alcohol. One is much more likely to be arrested if they are outside drinking in a bar or sports stadium than if they were at home. Furthermore, these individuals may also be breaking other liquor laws due to their privilege. An example of this would be illegally selling their legally obtained alcohol to minors.

The last assumption is that the assignment between the treatment and control group must be as good as random. At first glance this assumption holds, as individual are not able to control how old they are. Meaning that the estimated reduced form and first stage are consistent with the true reduced form and first stage. However, the first stage does not meet this assumption for the reasons mentioned in section 3 which include: desirability bias, measurement error and oversampling.

Even with the assumption that all of our criteria for the delta method are met, our economic intuition tells us that 866.46 Arrests per 10,000 person-years is simply too much of a difference from 82.73 which is the estimate that we derived utilizing the regression discontinuity method.

References

1. Carpenter, C., & Dobkin, C. (2015). The Minimum Legal Drinking Age and Crime. *Review of Economics and Statistics*, 97(2), 521-524. doi:10.1162/rest_a_00489

Figure 1: Various Bin-Widths for Drinking Profile

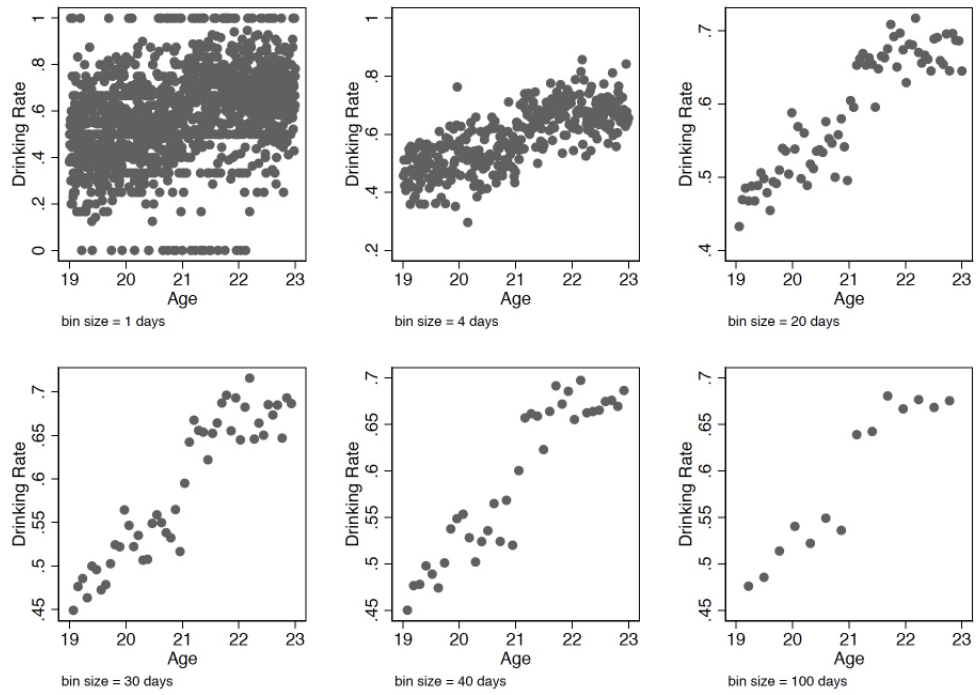


Figure 2: Various Bandwidths for Drinking Profile

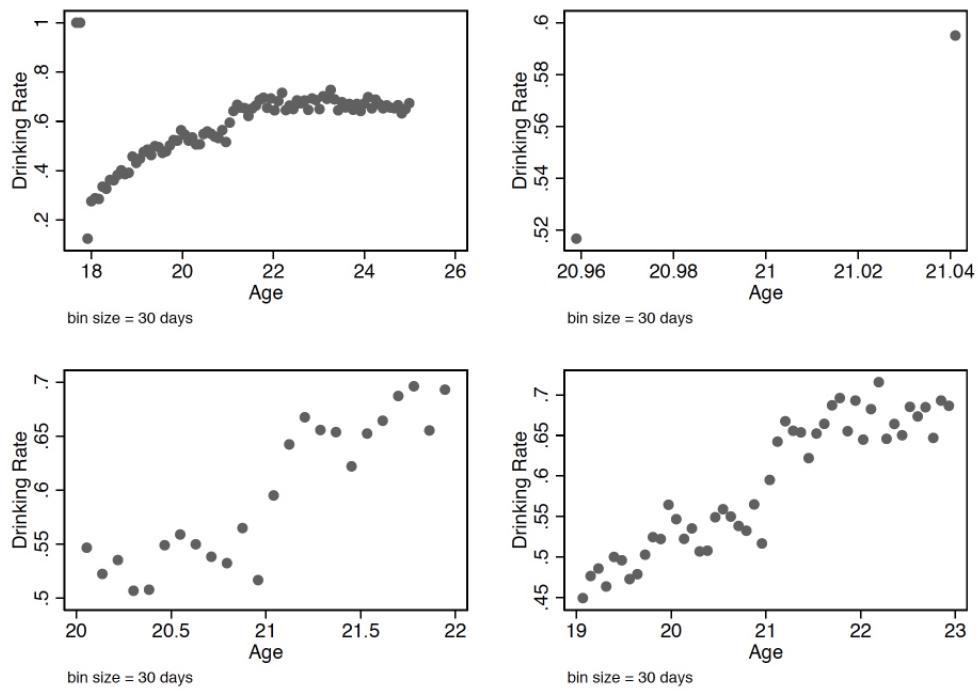


Figure 3: Various Drinking Rate Ranges for Drinking Profile

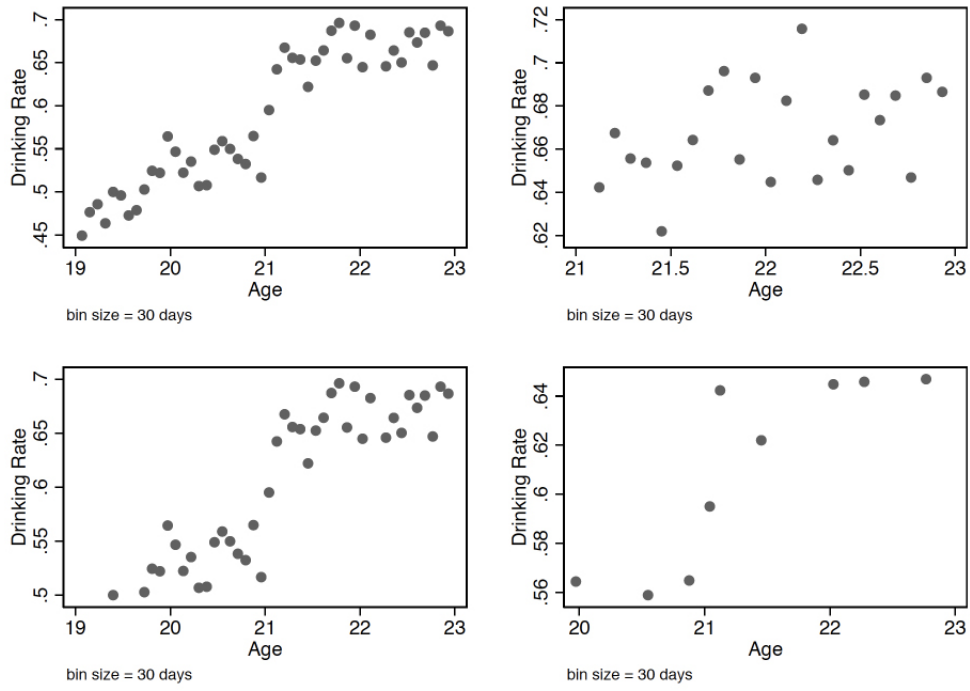


Figure 4: Various Polynomial Orders for Drinking Profile

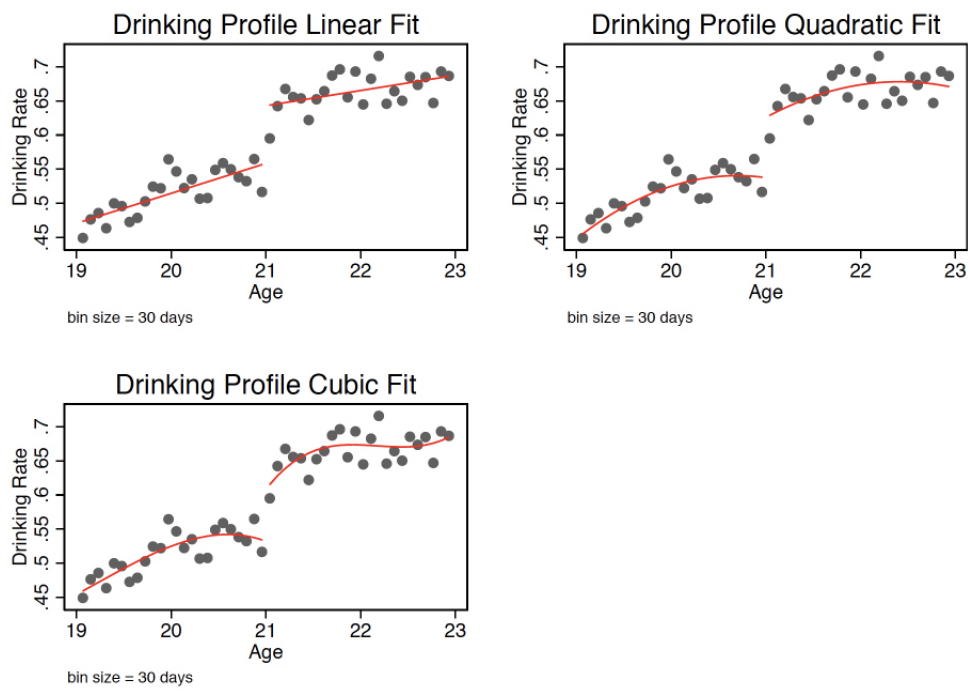


Figure 5: Age Profile of Drinking

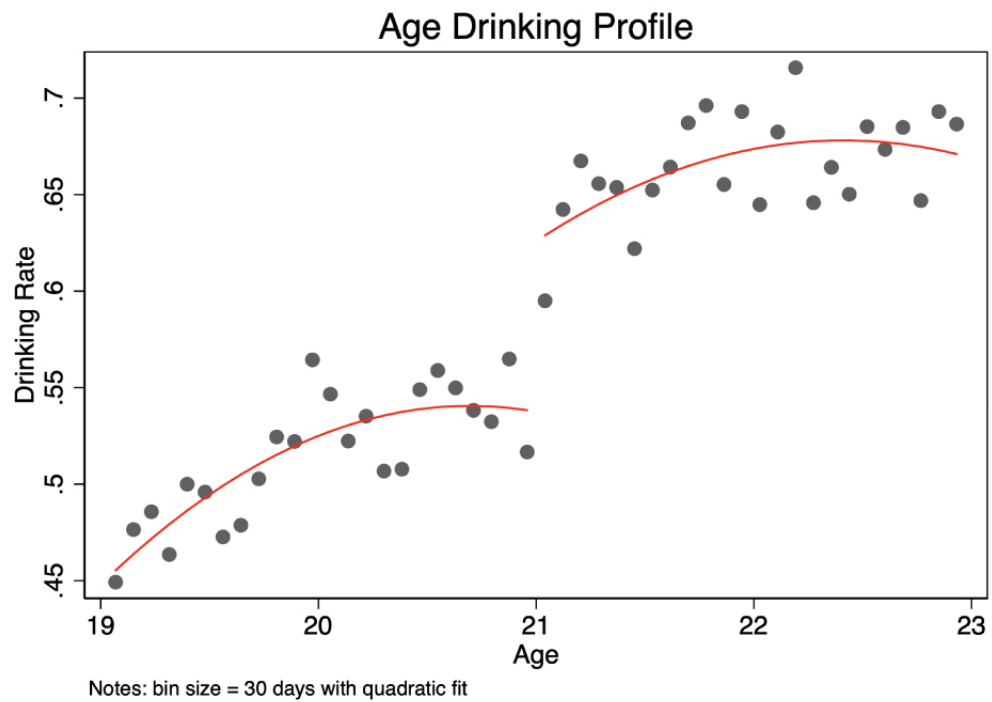


Figure 6: Age Profile of Arrests of the Combined Subcategories

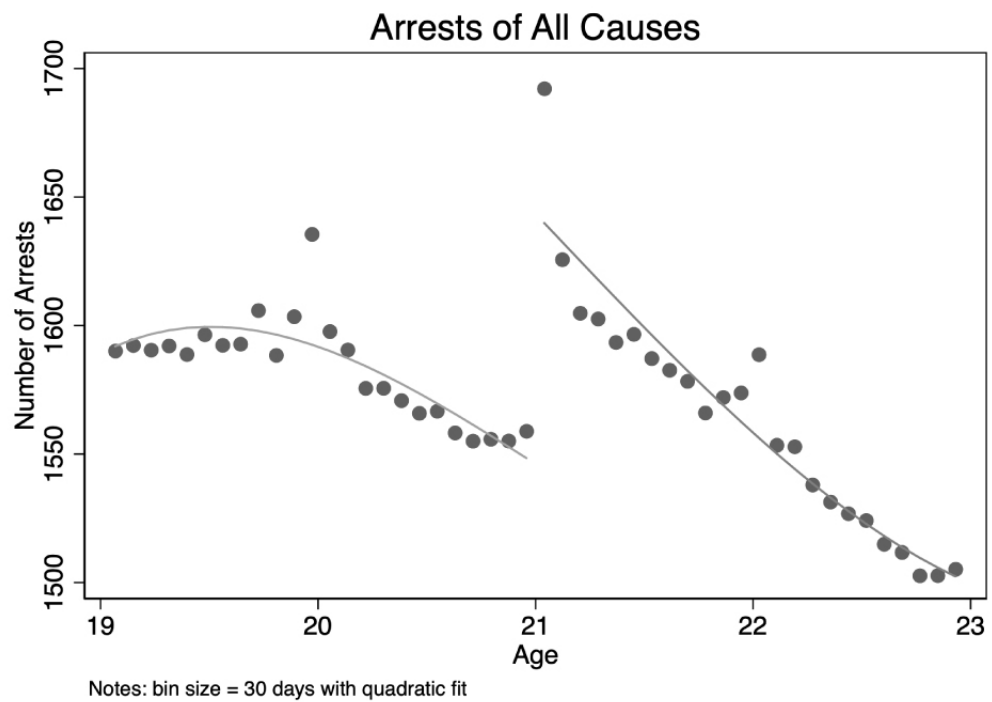


Figure 7: Age Profile of Subcategories of Arrests

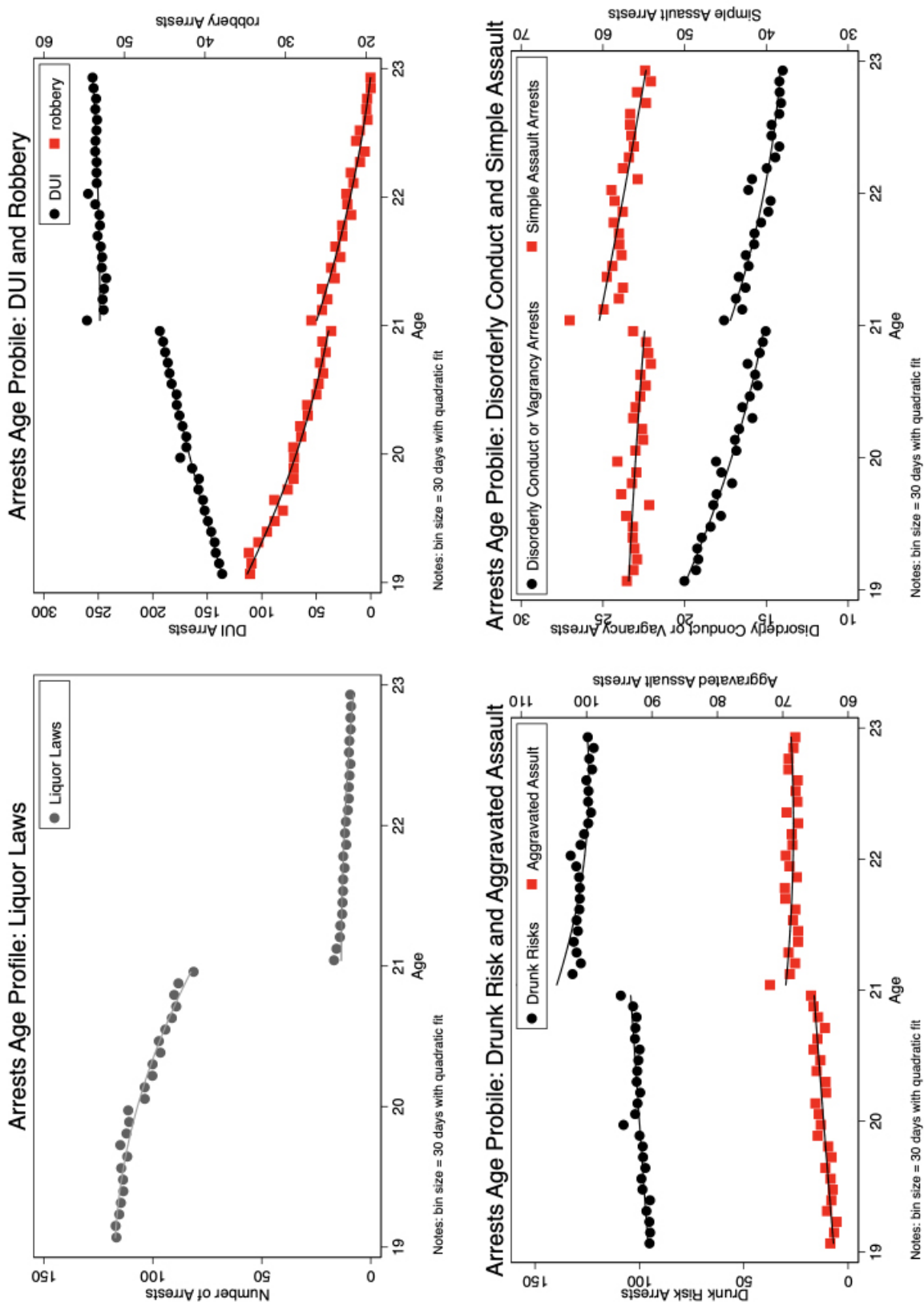


Table 1: Balance Table for Demographic Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Uninsured	HS Diploma	Hispanic	White	Black	Employed	Married	Working LW	Attends School	Male
MLDA	-0.001 (0.980)	0.033 (0.054)	-0.004 (0.844)	0.014 (0.531)	-0.024 (0.122)	0.028 (0.187)	-0.025 (0.107)	0.031 (0.143)	0.001 (0.927)	0.024 (0.276)
AgeCentered	-0.001 (0.968)	-0.064 (0.032)	0.002 (0.943)	-0.008 (0.831)	0.017 (0.516)	-0.013 (0.711)	0.067 (0.005)	-0.015 (0.680)	-0.061 (0.038)	-0.073 (0.046)
AgeCenteredxMLDA	-0.014 (0.770)	0.086 (0.030)	-0.023 (0.593)	0.013 (0.794)	0.009 (0.797)	0.058 (0.234)	-0.046 (0.183)	0.058 (0.238)	0.041 (0.285)	0.101 (0.045)
AgeCentered_squared	-0.014 (0.393)	-0.044 (0.003)	0.001 (0.943)	-0.000 (0.991)	0.004 (0.737)	-0.036 (0.040)	0.008 (0.461)	-0.037 (0.039)	-0.002 (0.919)	-0.025 (0.159)
AgeCentered_squaredxMLDA	0.026 (0.252)	0.034 (0.077)	0.009 (0.665)	-0.003 (0.897)	-0.015 (0.378)	0.031 (0.193)	0.008 (0.638)	0.032 (0.176)	-0.008 (0.669)	0.014 (0.565)
Constant	0.309 (0.000)	0.793 (0.000)	0.242 (0.000)	0.554 (0.000)	0.16 (0.000)	0.618 (0.000)	0.158 (0.000)	0.618 (0.000)	0.165 (0.000)	0.412 (0.000)
Observations	18,824	18,824	18,824	18,824	18,824	18,824	18,824	18,824	18,824	18,824
R-squared	0.001	0.003	0.000	0.000	0.000	0.014	0.019	0.014	0.019	0.000

Robust p-value in parentheses

Table 2: Regression Estimates of Change in Drinking Rate at 21 Years old

	(1)	(1b)	(2)	(2b)	(3)	(3b)
VARIABLES	Drinks Alcohol	Drinks Alcohol	Drinks Alcohol	Drinks Alcohol	Drinks Alcohol	Drinks Alcohol
MLDA	0.086*** (0.014)	0.086*** (0.014)	0.092*** (0.022)	0.091*** (0.022)	0.081*** (0.029)	0.078*** (0.029)
AgeCentered	0.044*** (0.009)	0.044*** (0.009)	-0.024 (0.037)	-0.024 (0.037)	-0.051 (0.093)	-0.051 (0.093)
AgeCenteredxMLDA	-0.024* (0.012)	-0.024* (0.012)	0.094* (0.050)	0.095* (0.050)	0.214* (0.125)	0.222* (0.126)
AgeCentered_squared			-0.034* (0.018)	-0.034* (0.018)	-0.068 (0.107)	-0.068 (0.107)
AgeCentered_squaredxMLDA			0.009 (0.024)	0.009 (0.024)	-0.072 (0.144)	-0.080 (0.146)
AgeCentered_cubed					-0.011 (0.035)	-0.011 (0.035)
AgeCentered_cubedxMLDA					0.050 (0.047)	0.052 (0.048)
Birthday		0.002 (0.084)		0.020 (0.083)		0.037 (0.084)
Constant	0.559*** (0.010)	0.559*** (0.010)	0.536*** (0.016)	0.536*** (0.016)	0.532*** (0.021)	0.532*** (0.021)
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

Table 3: Regression Estimates of Change in Arrests at 21 Years Old Not Accounting For Celebration Effect

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	DUI	Liquor Laws	Robbery	Aggravated Assault	Simple Assault	Drunk Risk	Disorderly Conduct
MLDA	105.791*** (12.003)	54.605*** (3.235)	-64.899*** (0.832)	1.798*** (0.401)	4.372*** (0.788)	5.772*** (0.936)	36.226*** (4.188)	2.205*** (0.429)
AgeCentered	-72.782*** (9.720)	24.987*** (2.449)	-33.608*** (1.931)	-2.901*** (0.704)	1.446 (1.144)	-1.247 (1.081)	4.224* (2.175)	-1.505*** (0.601)
AgeCenteredxMLDA	-38.929 (25.836)	-26.153*** (6.959)	26.373*** (1.997)	-2.341** (0.960)	-3.549** (1.795)	-2.259 (2.067)	-25.310*** (8.593)	-1.235 (0.957)
AgeCentered_squared	-24.002*** (5.094)	-2.507** (1.254)	-7.736*** (0.948)	1.232*** (0.360)	-0.060 (0.568)	-0.115 (0.541)	-0.078 (1.107)	0.509* (0.289)
AgeCentered_squaredxMLDA	43.471*** (11.494)	4.959 (3.095)	9.629*** (0.977)	-0.364 (0.473)	0.932 (0.862)	0.391 (0.954)	6.920* (3.683)	0.067 (0.443)
Constant	1,542.960*** (3.282)	194.210*** (0.894)	81.164*** (0.799)	24.546*** (0.269)	65.266*** (0.471)	54.862*** (0.434)	104.358*** (0.832)	15.116*** (0.250)
Observations	1,461	1,461	1,461	1,461	1,461	1,461	1,461	1,461
R-squared	0.410	0.937	0.990	0.719	0.236	0.101	0.615	0.320

Robust Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Regression Estimates of Change in Arrests at 21 Years Old Accounting For Celebration Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	All	DUI	Liquor Laws	Robbery	Aggravated Assault	Simple Assault	Drunk Risk	Disorderly Conduct
MLDA	82.729*** (5.249)	48.632*** (1.578)	-65.345*** (0.845)	1.532*** (0.464)	3.199*** (0.776)	4.165*** (0.723)	27.981*** (1.436)	2.021*** (0.399)
AgeCentered	-72.782*** (9.724)	24.987*** (2.449)	-33.608*** (1.932)	-2.901*** (0.704)	1.446 (1.145)	-1.247 (1.081)	4.224* (2.176)	-1.505** (0.601)
AgeCenteredxMLDA	5.831 (14.657)	-14.561*** (4.196)	27.239*** (2.017)	-1.824* (1.053)	-1.273 (1.762)	0.859 (1.697)	-9.307** (3.618)	-0.878 (0.900)
AgeCentered_squared	-24.002*** (5.096)	-2.507** (1.254)	-7.736*** (0.948)	1.232*** (0.360)	-0.060 (0.569)	-0.115 (0.542)	-0.078 (1.108)	0.509* (0.289)
AgeCentered_squaredxMLDA	25.110*** (7.407)	0.203 (2.080)	9.274*** (0.984)	-0.576 (0.503)	-0.002 (0.849)	-0.888 (0.818)	0.355 (1.769)	-0.079 (0.421)
Birthday	66.055** (27.826)	17.108** (7.470)	1.279*** (0.434)	0.762 (0.546)	3.358** (1.332)	4.601** (1.916)	23.616** (9.957)	0.527 (0.790)
Constant	1,542.960*** (3.283)	194.210*** (0.895)	81.164*** (0.799)	24.546*** (0.269)	65.266*** (0.471)	54.862*** (0.435)	104.358*** (0.832)	15.116*** (0.250)
Observations	1,461	1,461	1,461	1,461	1,461	1,461	1,461	1,461
R-squared	0.432	0.939	0.990	0.719	0.241	0.115	0.638	0.321

Robust Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: IV Estimates of the Effect of Drinking Alcohol for Arrests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	All	DUI	Liquor Laws	Robbery	Aggravated Assault	Simple Assault	Drunk Risk	Disorderly Conduct
IV Estimate	866.46***	532.07***	-733.64***	17.48***	34.99***	46.66***	308.43***	22.35***
SE	(188.35)	(112.89)	(154.55)	(5.79)	(10.49)	(11.98)	(17.83)	(5.01)

Robust Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1