The Contrast Between Random Control Trial

And Non Experimental Analysis

Devin Maya

6/04/2023

ABSTRACT

This paper provides an in-depth analysis of the Minimum Legal Drinking Age's (MLDA) causal impact on alcohol consumption and associated crime rates. Drawing upon data from the National Health Interview Sample Adult Files from 1997-2007 and a specific dataset detailing arrest rate. The regression discontinuity method is implemented to understand changes in drinking habits around the age threshold of 21. The analysis suggests a 10-percentage point increase in alcohol consumption as individuals cross the MLDA. Additionally, there is an uptick in arrest rates post the legal drinking age, which could be directly attributed to the increased alcohol consumption, or indirectly, through greater exposure to venues like clubs and bars. The paper further delves into a methodological exploration, deploying an instrumental variable (IV) approach to compute the MLDA's effect on per-person drinking and arrest rates. The fulfillment of critical IV assumptions is questioned in this context. This investigation into MLDA's effects provides crucial insights with substantial implications for policymaking in public health and safety spheres.

Introduction

In the domain of public health policy, the establishment of the Minimum Legal Drinking Age serves as a significant topic in shaping societal behavior, particularly relating to alcohol consumption and its subsequent consequences. This in-depth paper delves into the effects of the MLDA on alcohol consumption and crime, aiming to highlight the complex interplay between legal age boundaries, personal behavior, and societal outcomes. The findings from this research hold substantial implications for potential policymakers, public health practitioners, and society in general.

Drawing upon the data from the National Health Interview Sample Adult Files (1997-2007) and corresponding arrest records. The data set meticulously records pertinent variables such as demographic information, employment status, and specificities related to drinking habits and crime incidents, making them good sources for comprehensive analysis. The empirical methodology utilized in this research adopts advanced econometric methods. These methods include age-specific profiling of drinking and arrest patterns, alongside various regression models tailored to the research questions at hand. Care is taken in selecting the suitable support for age and the appropriate polynomial order for the regression analyses, ensuring a high level of precision in capturing the relationship between MLDA, alcohol consumption, and crime rates.

A key finding reveals a discernible discontinuity at the age threshold of 21, corresponding to a rise in the proportion of individuals reporting alcohol consumption. The results display various statistical findings and interpretations. These include a detailed examination of age profiles in relation to drinking habits and crime, the presentation of regression estimates with different specifications, and a close look at the sharp changes in other covariates at age 21. Our study also uncovers an interesting observation with an increase of approximately 10 percentage points in alcohol consumption as individuals cross the MLDA. While the MLDA's effect on per person drinking and arrest rates is estimated through an IV approach, certain assumptions that must be satisfied for this methodology to be sound come into question in this context. Additionally, the analysis uncovers intriguing correlations between MLDA, alcohol

consumption, and crime rates, inviting deeper exploration into the wider societal impacts of legal drinking age policies.

Beyond simple correlations, the study employs an instrumental variables approach in combination with regression discontinuity design at the MLDA threshold in order to examine the effects of the MLDA on arrests per person drinking, providing a refined understanding of the relationship between alcohol consumption and crime. This technique allows for navigating potential confounding factors and isolating the specific influence of the MLDA on crime incidents.

This research emphasizes the significant role of age-related drinking laws in public health and safety. The results illuminate the necessity for ongoing policy evaluation and fine-tuning, informed by a comprehensive understanding of the implications of the MLDA. The paper encapsulates the findings and implications, while also addressing the robustness of the methods and assumptions. The insights drawn from this investigation carry significant relevance for public health and safety policy considerations. Overall providing a rigorous data-driven examination of the MLDA's effects on alcohol consumption and crime rates and examining the validity of implementing such a design.

Data

The research utilizes two datasets to estimate the effects of the Minimum Legal Drinking Age on alcohol consumption and crime rates. The data is derived from the National Health Interview Sample Adult Files (NHIS) spanning from 1997 to 2007, and a dataset named "Arrest.csv" detailing arrest rates per 10,000 by age, including breakdowns by cause. It has been extensively used in economic, health, and social

behavior research. This dataset is particularly valuable due to its longitudinal nature, allowing the tracking of individual behaviors and outcomes over time. For a comprehensive understanding of the data, you may want to refer to Carpenter and Dobkin (2015), who provide a detailed analysis using the same dataset to study the effects of alcohol on mortality Carpenter, C., & Dobkin, C. (2015). The Minimum Legal Drinking Age and Morbidity in the United States. Review of Economics and Statistics, 97(2), 521–524.

The NHIS dataset is part of a series of cross-sectional household interview surveys designed to collect data on the health of the United States population. Conducted annually by the National Center for Health Statistics (NCHS), a part of the Centers for Disease Control and Prevention (CDC), the data is a nationally representative sample of the civilian, non-institutionalized population of the United States.

In this analysis a subset of the NHIS adult files is used while comprising several key variables to perform the analysis. Some of the variables within the NHIS dataset include indicators of educational attainment, ethnicity, insurance status, employment status, marital status, work or school participation, and gender. Specific variables related to alcohol consumption are Days to 21st birthday, reports they drink alcohol, Age, and Percent of days on which they report drinking. Each row in the dataset is a unique record for an individual, enabling person-level analysis.

The Arrest.csv dataset is a comprehensive collection of age-specific arrest rates, offering a resource for studying the correlation between the MLDA, alcohol consumption, and crime rates. The dataset, Arrest.csv is a comprehensive compilation

of arrest rates per 10,000 individuals by age. This dataset includes categories for total arrests as well as those specific to alcohol-related offenses, broken down by cause. Each row represents the arrest rates for a specific age, including some notable variables such as age, all arrests, drunk possibly risk to self, driving under the influence, violation of liquor law, disorderly conduct or vagrancy, robbery, simple assault, and aggravated assault.

Together, these datasets provide a broad spectrum of information, allowing us to explore the complex interplay between the MLDA, alcohol consumption, and crime rates. In ensuring a robust analysis, the use of such data has been reviewed thoroughly verifying their validity and addressing any potential discrepancies or missing values.

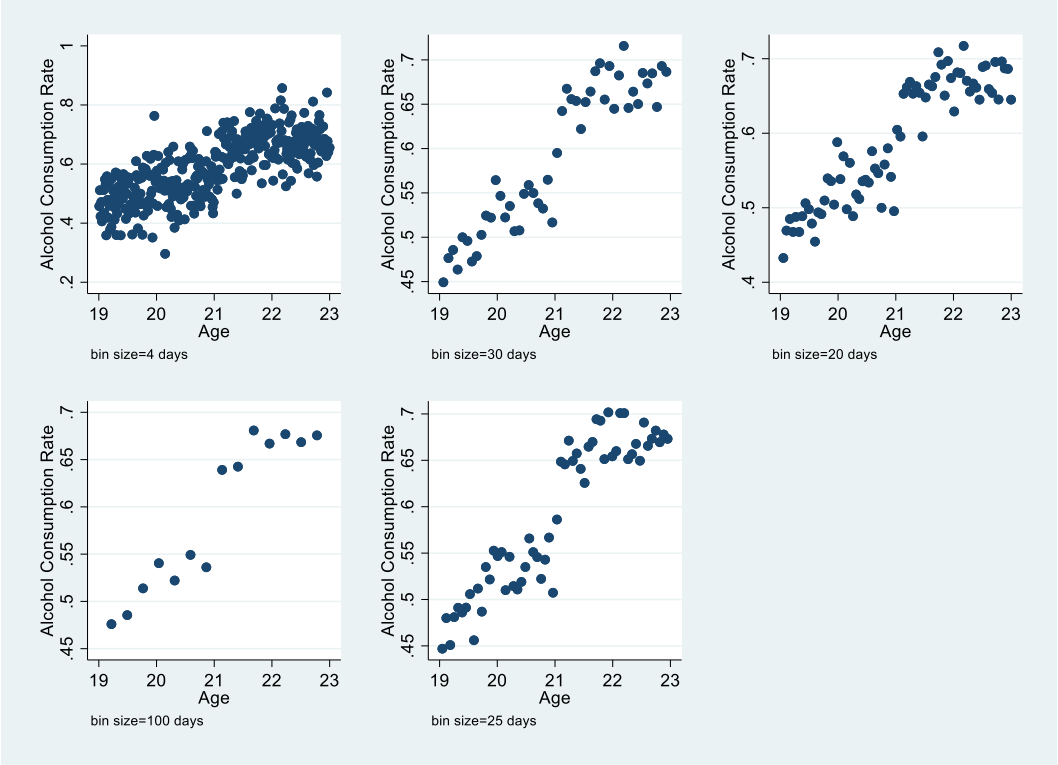
Methods

The primary econometric method utilized in this research is the regression discontinuity design. This approach leverages the sharp discontinuity in legal access to alcohol at the age of 21, estimating the causal effect of the Minimum Legal Drinking Age (MLDA) on the variables of interest. This method essentially compares individuals who are just under 21 with those just over 21, assuming that, aside from legal access to alcohol, the two groups are otherwise similar.

Regarding the variables used in the graphical results, specific selection ranges are essential. The age variable spans from the minimum to the maximum ages in the dataset, focusing particularly on the legal drinking age threshold of 21.The implementation of polynomial regression models of various orders is used for estimating the relationship between age, drinking behavior, and crime rates. Guided by preliminary

data exploration and graphical results combined with statistics, a quadratic model offers the optimal balance between model fit and overfitting prevention.

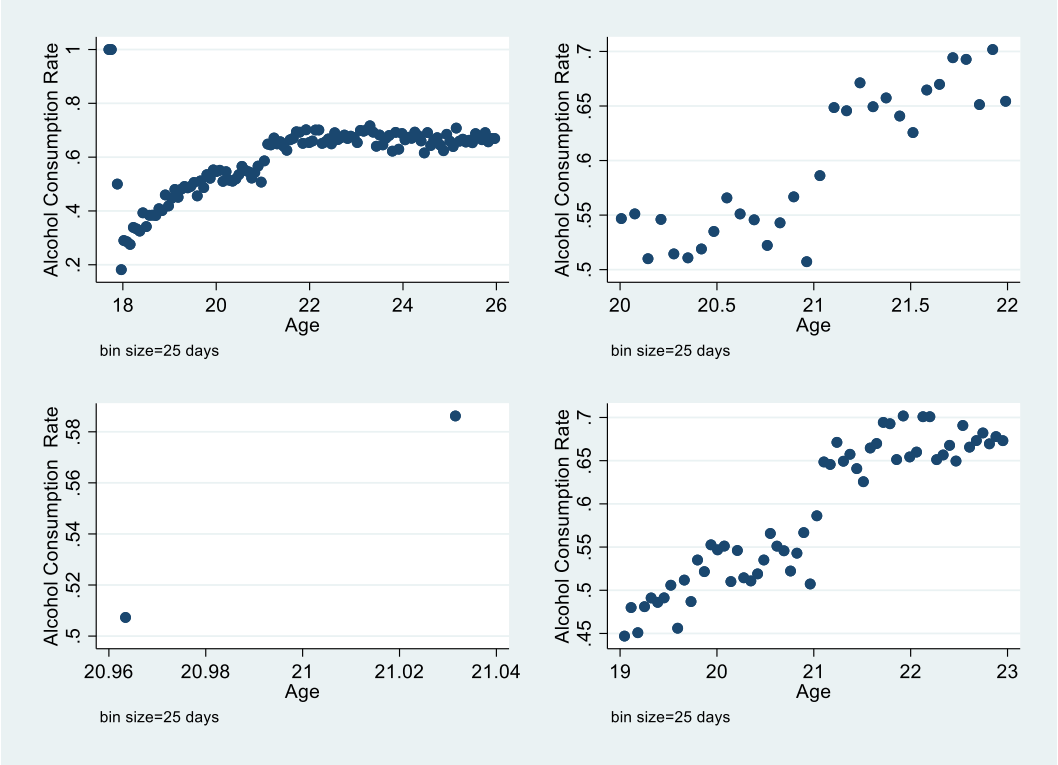
The study uses regression discontinuity design, where the selected bin width and bandwidth for age are centered around the legal drinking age threshold of 21 years. A critical factor was the selection of bin width, which impacts noise reduction and data bias minimization. An optimal choice of 25-day bin width was found to ensure clear observations of discontinuities in the data while balancing noise reduction and the accurate representation of the true relationship.



Above is a panel of figures which display different bin width ranges and their given results. This choice of 25 days for the bin width enables the ability to observe discontinuities in the data with clarity while balancing the trade-off between noise reduction and representation of the true relationship being examined. Also determining

an appropriate range of age to be between 19 and 23 years old which ensures the inclusion of representative observations for the RD Design.

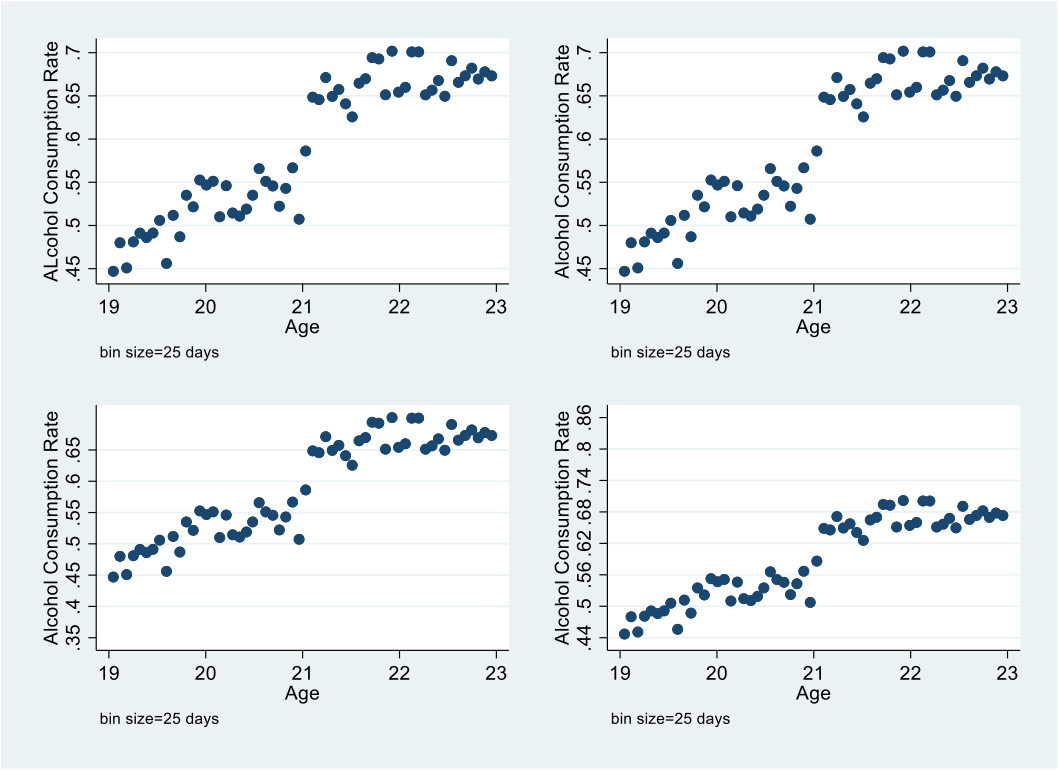
Determining an appropriate age range between 19 and 23 years old ensured the inclusion of representative observations for the Regression Discontinuity (RD) Design, displaying a clear discontinuity while avoiding the exclusion of relevant observations. For the dependent variable, in this case, the Alcohol Consumption Rate, selection of an appropriate range is also crucial to prevent over- or understatement of the true discontinuity.



This bandwidth selection displays a clear discontinuity while avoiding the exclusion of relevant observations which are shown on the x axis range of the figures. Selection for the appropriate range for the dependent variable is also crucial. In this case the Alcohol Consumption Rate must be selected in a way to prevent the overstatement or understatement of the true discontinuity. Ultimately, the decision came

to a y-range between .45-.75 for the Alcohol Consumption Rate allowing to effectively illustrate the discontinuity occurring at the legal drinking age.

A y-range between .45-.75 for the Alcohol Consumption Rate was selected to effectively illustrate the discontinuity occurring at the legal drinking age.



The figures display various ranges and adjustments to the y range value displaying the reasoning for the selected thresholds. Two key regression equations are employed. One for the relationship between age and alcohol consumption, and one for the relationship between age and crime rates. The specifications include various control variables, including various characteristics and variables to account for the effects at the age of 21. In terms of polynomial order, the objective is to balance the flexibility of the model with the risk of overfitting. A low-degree polynomial such as a linear or quadratic was used to capture the relationship without adding undue complexity.

The analysis relies on two primary regression equations. One for the relationship between age and alcohol consumption the First Stage , and one for the relationship between age and crime rates the Reduced Form . For the FS model, the equation is:

$$drink_i = \varphi_0 + \varphi_1 Post_i + g(age_i) + \varepsilon_i$$

Here, $drink_i$ is the outcome variable, indicating whether individual i drinks or not. φ_0 is the intercept term in the regression model, φ_1 is the coefficient for $Post_i$, a binary variable indicating whether individual i is over the MLDA of 21 years, and $g(age_i)$ is a function of age, capturing the underlying relationship between age and drinking behavior. The purpose of this first-stage regression is to estimate the causal effect of reaching the MLDA on drinking behavior.

For the RF model, the equation is:

$$arrest_i = \rho_0 + \rho_1 Post_i + g(age_i) + \varepsilon_i$$

Now the only difference is that $arrest_i$ is the outcome variable, representing the crime rates of individual i. everything else remains the same when comparing to the First Stage. The goal of this reduced form regression is to estimate the overall effect of the MLDA policy on the mortality rate.

These methodological decisions ensure robust analysis and help to avoid violation of the assumptions of regression analysis. The validity of these methods, the choice of regression specifications, and the adherence to necessary assumptions are all crucial when analyzing the results.

The goal of this reduced form regression is to estimate the overall effect of the MLDA policy on the mortality rate. The coefficient ρ1 captures the average change in the mortality rate at the threshold of 21 years, which represents the combined effect of legal access to alcohol and the resulting change in drinking behavior on mortality. These overall decisions also helped in ensuring the assumptions of regression analysis were not violated. The appropriateness of these methods, the choice of regression specifications, and the validity of the necessary assumptions are all crucial and relevant when analyzing the results.

Results

The Results section of this paper delves into a comprehensive exploration of the impact of the Minimum Legal Drinking Age on alcohol consumption and related arrests. This is examined through a variety of figures and regression tables, each of which offers a unique perspective on the relationship between the MLDA, alcohol consumption, and crime rates.

Figure 1

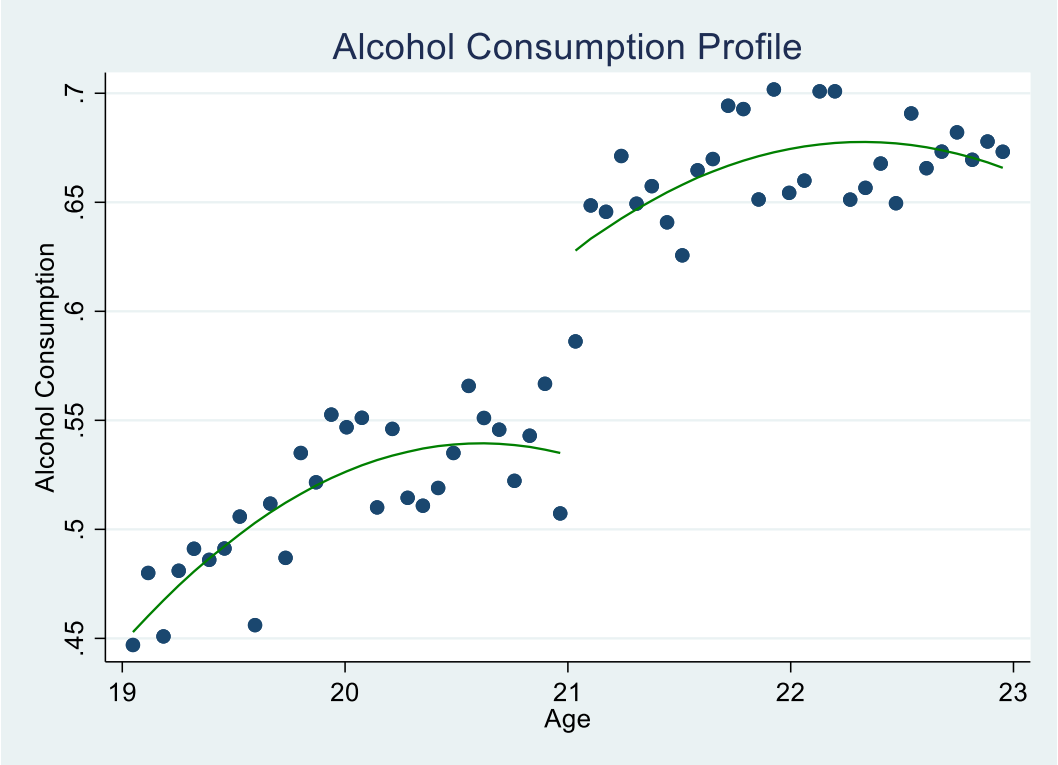


Figure 1 provides a graphical representation of the relationship between age and alcohol consumption, presenting the transition in drinking behavior around the legal drinking age. This visual representation is further quantified in Table 1, which displays a regression table highlighting the effect of the Minimum Legal Drinking Age (MLDA) on alcohol consumption, under different model specifications and adjusting for potential birthday celebration effects. The point estimate for the MLDA effect on drinking, for instance, indicates a 8.7 percentage point increase in alcohol consumption when an individual crosses the age threshold of 21. This value, expressed in both linear and quadratic with no drastic discrepancy is statistically significant at the 1% significance

level, meaning there's strong evidence to suggest the MLDA has a significant effect on drinking behavior.

Table 1

VARIABLES	(1) Alcohol Consumption	(2) Alcohol Consumption	(3) Alcohol Consumption
Over21	0.087***	0.091***	0.081***
	(0.014)	(0.021)	(0.029)
(Age-21)	0.044***	-0.024	-0.051
	(0.009)	(0.036)	(0.090)
(Age-21)*Over21	-0.024**	0.097*	0.207*
	(0.012)	(0.049)	(0.124)
(Age-21)^2		-0.034*	-0.068
		(0.017)	(0.104)
(Age-21)^2*Over21		0.007	-0.062
		(0.024)	(0.144)
(Age-21)^3			-0.011
			(0.034)
(Age-21)^3*Over21			0.046
			(0.047)
Constant	0.559***	0.536***	0.532***
	(0.010)	(0.015)	(0.021)
Observations	18,801	18,801	18,801
R-squared	0.025	0.025	0.025

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next, to ensure the validity of the research design, the inclusion of a balance table (**Table 2**) across various demographic and socioeconomic characteristics is necessary. This table serves to confirm that there are no significant changes in these characteristics at the age cutoff, thereby strengthening the claim of a causal effect of MLDA on alcohol consumption and related crime rates. Specifically, the values for the

associated coefficients are similar in value as can be determined by the regression results where the linear model consisted of a value of 111.13 and the quadratic model consisted of 112.65. This comparison is clearly shown with the results below.

Table 2

```
. ** Linear fits
. reg drinks_alcohol over21 agec agec_post if age>=19 & age<23
```

Source	SS	df	MS	Number of obs	=	18,801
				F(3, 18797)	=	157.46
Model	111.134887	3	37.0449623	Prob > F	=	0.0000
Residual	4422.19781	18,797	.235260829	R-squared	=	0.0245
				Adj R-squared	=	0.0244
Total	4533.3327	18,800	.241134718	Root MSE	=	.48504

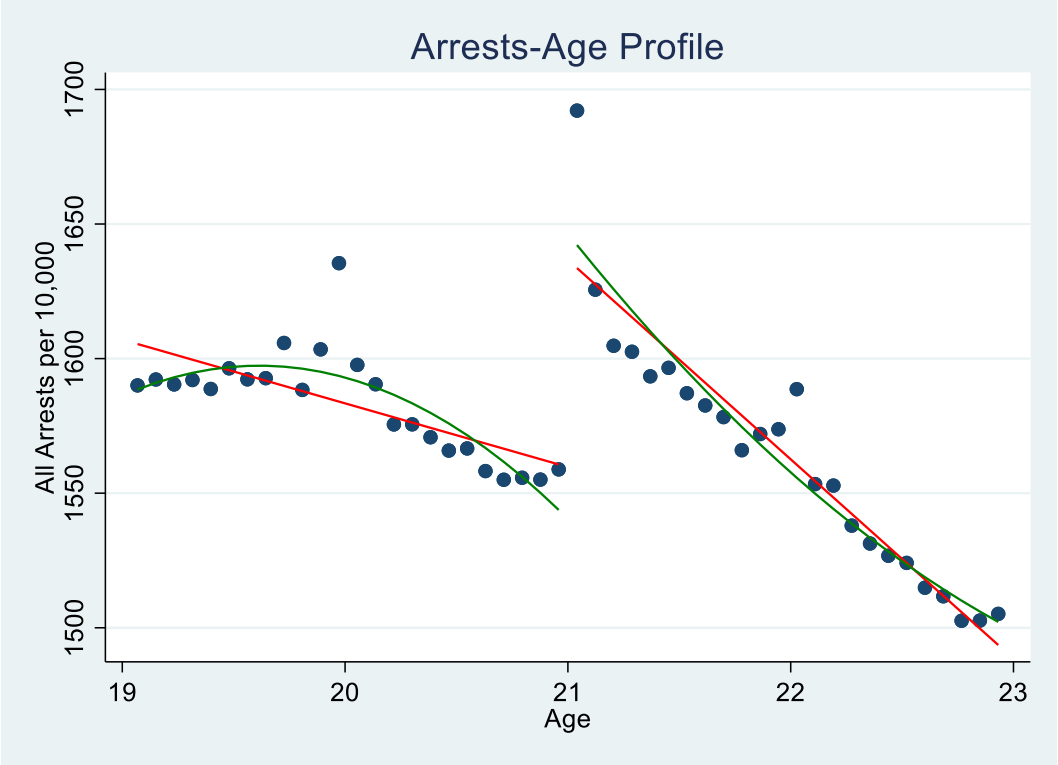
drinks_alc~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
over21	.0865785	.0141262	6.13	0.000	.0588899	.1142671
agec	.0438739	.0089661	4.89	0.000	.0262995	.0614483
agec_post	-.0243453	.0123268	-1.97	0.048	-.0485069	-.0001838
_cons	.5585858	.0101653	54.95	0.000	.5386608	.5785107

```
. ** Quadratic fits
. reg drinks_alcohol over21 agec agec_sq agec_post agec_sq_post if age>=19 & age<23
```

Source	SS	df	MS	Number of obs	=	18,801
				F(5, 18795)	=	95.79
Model	112.649416	5	22.5298833	Prob > F	=	0.0000
Residual	4420.68328	18,795	.235205282	R-squared	=	0.0248
				Adj R-squared	=	0.0246
Total	4533.3327	18,800	.241134718	Root MSE	=	.48498

drinks_alc~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
over21	.0908001	.0213173	4.26	0.000	.0490162	.132584
agec	-.0236293	.0358255	-0.66	0.510	-.0938505	.0465918
agec_sq	-.0340016	.0174712	-1.95	0.052	-.0682468	.0002436
agec_post	.096858	.0494216	1.96	0.050	-.0000128	.1937287
agec_sq_post	.0072699	.0239734	0.30	0.762	-.0397201	.05426
_cons	.536271	.0153226	35.00	0.000	.5062374	.5663047

Figure 2



Observing the results for Figure 2 this is illustrating the age profile of arrests due to all causes. As shown at the threshold there is a significant increase at the threshold for all arrests on the given age profile. This therefore implies that the MLDA does have a significant impact on the overall arrests. Now figures examining the sub causes of arrests included within the data set around the threshold will be provided.

Figure 3

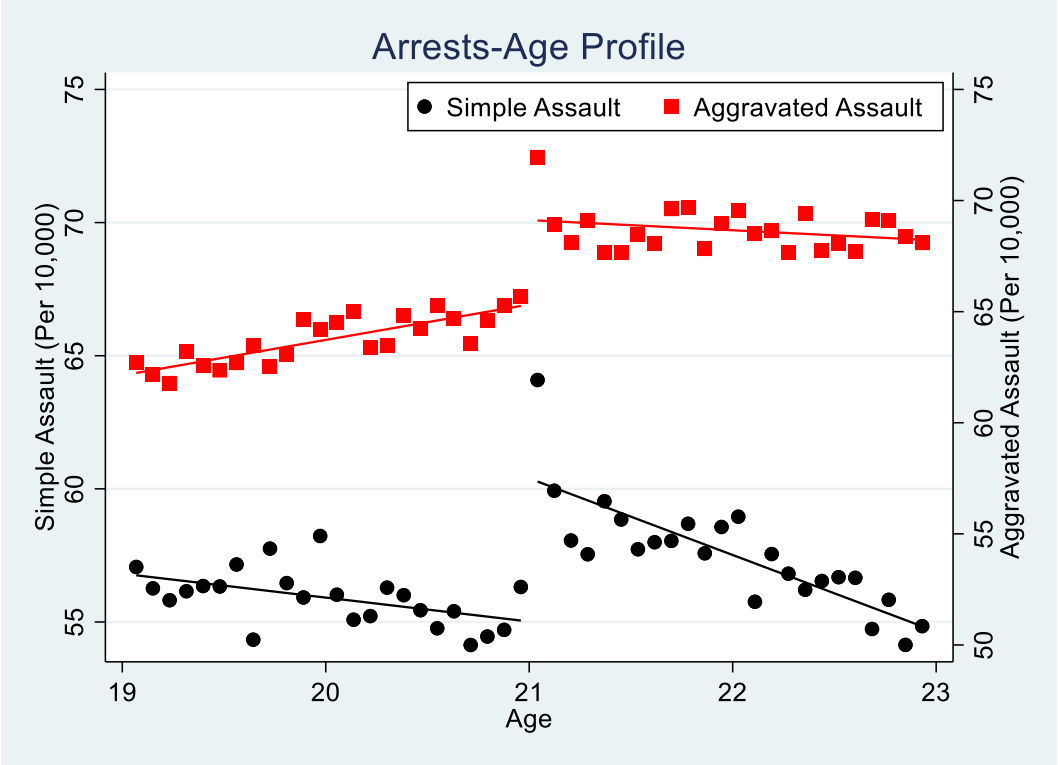
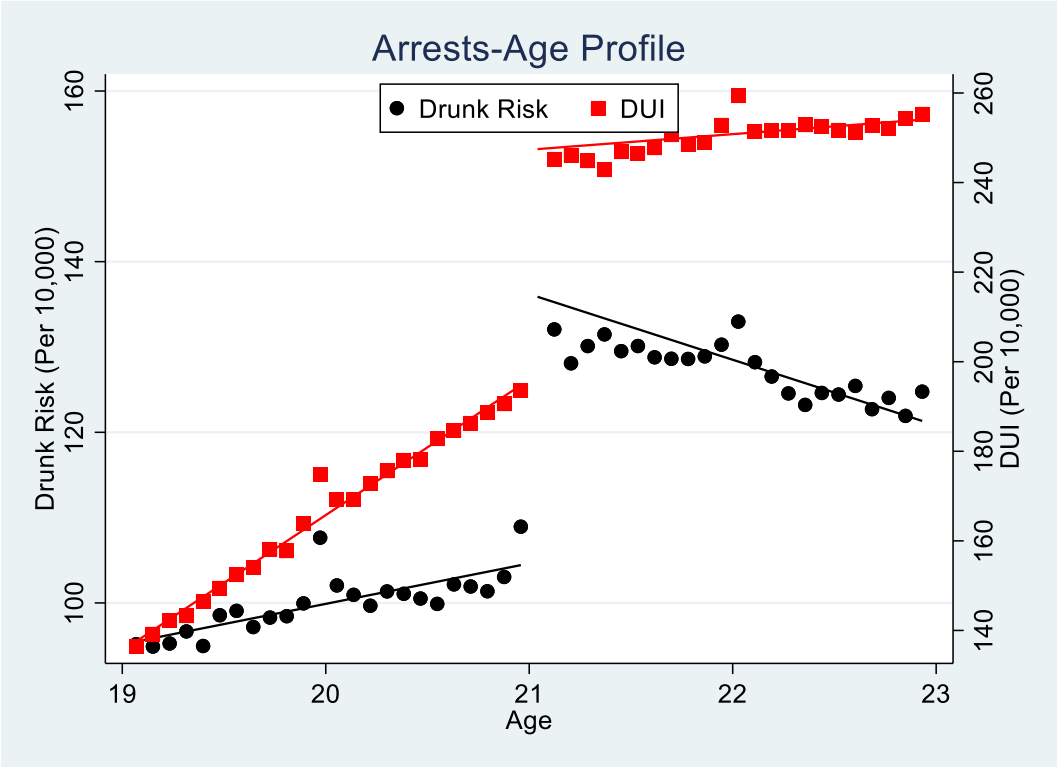


Figure 3 displays the sub causes of arrest that are simple assault and aggravated assault. The relationship highlights an increasing value at the threshold age of 21 years old for both. Specifically simple assaults drop significantly the older the induvial gets as shown by the decreasing slope after the threshold. Producing the same figures for the other sub causes of arrests will provide insight into the relationships within the data when using regression discontinuity.

Figure 4



Individuals who were included in the data for the analyzation of being a drunk risk which is possibly a risk to self and DUI both display an increase at the threshold. For both sub causes it is evident that the MLDA produces a discontinuity in which the overall values increase significantly. After the MLDA it is shown that Drunk risk goes down while DUI's continue to increase at a minimal amount which is observable through their increasing and decreasing slopes.

Figure 5

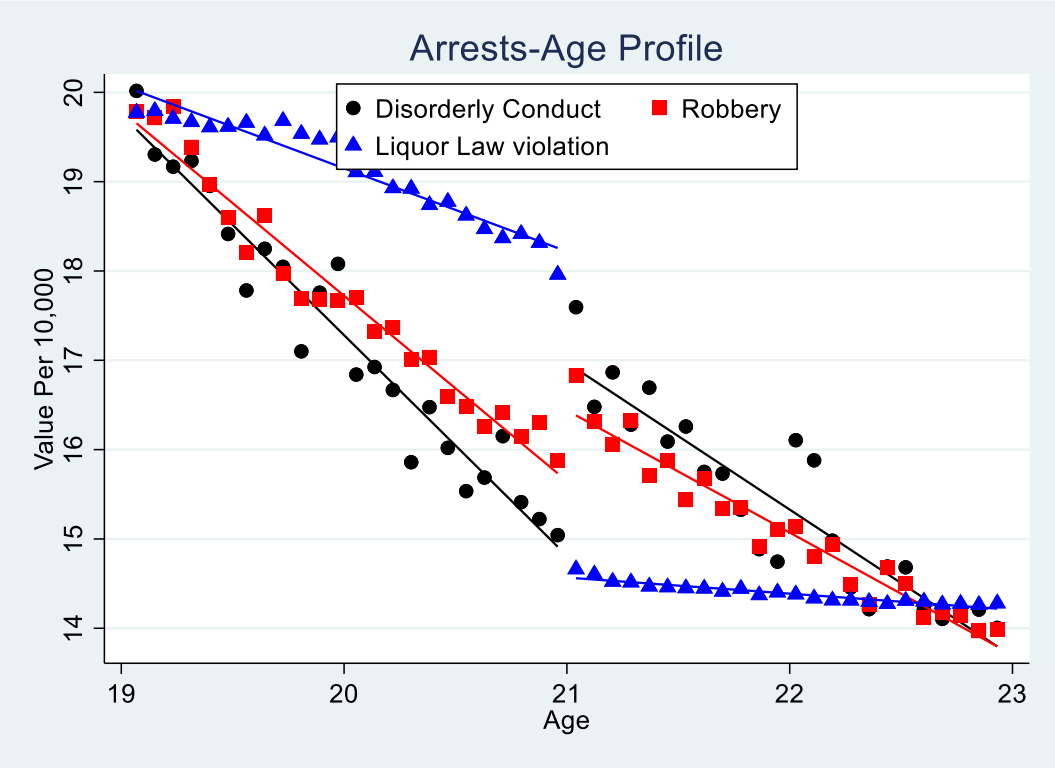


Figure 5 includes now 3 sub causes of arrests which include robbery, liquor law violations, and disorderly conduct. It is observed that the three sub causes of arrest have a substantial decrease effect due to the MLDA. This is a relationship that can be interpreted for a decrease in liquor law violations since the induvial can now drink legally thus the biggest decrease is without a doubt displayed when observing the discontinuity for liquor laws being violated.

Table 3

VARIABLES	(1)	(2)	(3)	(4)
Aggravated Assault	0.095***	0.035	0.035	0.023
	(0.025)	(0.024)	(0.024)	(0.024)
Drunk Risk		0.120***	0.119***	0.080***
		(0.010)	(0.010)	(0.011)
Robbery		0.039	0.035	0.033
		(0.041)	(0.040)	(0.040)
Liquor law Violation		0.025	0.022	0.007
		(0.024)	(0.023)	(0.022)
over21	5.192***	2.745	3.125*	-0.218
	(0.472)	(1.757)	(1.723)	(1.796)
agec	-1.165***	-0.866	-1.009*	-2.870***
	(0.286)	(1.354)	(0.547)	(0.671)
agec_post	-1.811***	0.509	-0.890*	0.517
	(0.403)	(1.677)	(0.532)	(0.597)
agec_sq		0.039		
		(0.555)		
agec_sq_post		-0.795		
		(0.772)		
Dui				0.066***
				(0.012)
Disorderly Conduct or Vagrancy				0.086*
				(0.049)
Constant	48.738***	37.037***	37.506***	29.401***
	(1.662)	(2.681)	(2.656)	(2.945)
Observations	1,460	1,460	1,460	1,460
R-squared	0.112	0.202	0.201	0.221

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 delivers a comprehensive insight into the relationship between the minimum legal drinking age and various subcategories of arrests. These subcategories

include variables such as aggravated assault, being a drunk risk, robbery, and liquor law violations.

The general effect of transitioning over the MLDA threshold on arrest rates (represented by the variable 'Over 21' varies across the models. In model (1), this effect is highly significant at the 1 percent significance level with a coefficient of 5.19. However, in model (4), the effect is not statistically significant, as the coefficient is -0.22 with a high standard error of 1.796.

When examining alcohol-related and non-alcohol-related crime rates separately, we observe divergent effects. Specifically, when considering alcohol-related crime rates, illustrated by variables such as 'Drunk Risk' and 'Liquor Law Violation', there is a statistically significant increase p value for model (3) on Drunk Risk in the associated arrest rates as individuals transition over the MLDA.

Conversely, the impact of crossing the MLDA on non-alcohol-related crime rates, represented by variables such as 'Aggravated Assault' and 'Robbery', appears less significant. For instance, in model (4), 'Aggravated Assault' and 'Robbery' are not statistically significant, suggesting the influence of MLDA on these crime rates is less pronounced.

Furthermore, newly introduced variables such as 'DUI' and 'Disorderly Conduct or Vagrancy' in model (4) show that specific criminal behaviors such as driving under influence also experience a significant increase for dui post the MLDA threshold.

These results provide a nuanced understanding of the MLDA's effect on crime, indicating that its impact varies based on the nature of the crime. While alcohol-related

crimes see a notable rise upon reaching the MLDA, the same cannot be broadly said for non-alcohol-related crimes.

Conclusion

In the conducted analysis, it was observed that there was a notable 10 percentage point increase in individuals reporting alcohol consumption upon reaching the minimum legal drinking age .This rise underlines the profound influence the MLDA policy holds over drinking behaviors.

Employing an Instrumental Variable approach, using the legal drinking age as an instrument for alcohol consumption, enabled the exploration of the causal effect of drinking on crime rates. The first stage does not equal 0, as was affirmed by the results presented in Figure 1, demonstrating the relevance of the instrument. Moreover, it's assumed that the instrument influences the outcome, crime rates, solely through its effect on alcohol consumption, thereby reinforcing the exclusion restriction assumption.

The IV approach suggested a positive relationship between alcohol consumption and criminal activities. The point estimate was 1155.49, and the standard error was 280.04. Both are significant, indicating a robust relationship. However, these IV estimates should be interpreted with caution given the assumptions required for their validity. The large magnitude of the point estimate might suggest an upward bias.

Even though the IV approach assumes that no other channels affect the outcome, scenarios that violate this assumption could create an upward bias in the IV estimate. For instance, individuals who drink before the MLDA are likely to increase their consumption post-MLDA since they can now legally purchase alcohol. Similarly, gaining access to certain venues or events at the age of 21 could influence arrest rates.

The IV estimate, in this context, likely represents a Local Average Treatment Effect (LATE), capturing the effect of alcohol consumption on crime rates for the compliers (those who change their drinking behavior at 21 due to the law) rather than the always-takers or never-takers. This implies that the LATE might not fully reflect the effects on the entire population.

All in all, the findings offer compelling evidence of the MLDA's impact on alcohol consumption and its consequential influence on crime rates. These results should be interpreted with careful consideration due to the inherent assumptions in the IV approach. This research contributes valuable insights into the societal impacts of alcohol policies, emphasizing the need for a critical analysis when considering both the direct and indirect effects accompanying the legal drinking age. Future research could aim to substantiate these findings and further delve into the potential mechanisms at work.