Utilizing the Regression Discontinuity Design to Demonstrate the Effect of the MDLA on Mortality

By: Soham Shewale

March 14, 2018

The minimum legal drinking age was put in place to curb the negative outcomes associated with alcohol consumption, but how effective is it? This study looks at the effect of alcohol on mortality through the regression discontinuity design. The data was taken from the National Health Interview Sample and National Vital Statistics and gives traits of people who complied to an interview. However, there is something different between people who drink alcohol and those who don't, so we use the instrumental variable to get an estimate of compliers. As we compare the results, we see that turning 21 increases alcohol consumption by 8.7% and increases mortality by 9.548 deaths per 100,000 individuals. Using these two estimates, we find that the instrumental variable estimate is 109.747 deaths per 100,000 individuals.

Introduction

Alcohol consumption can cause harm to society; people who consume a large amount of alcohol have the ability to harm not only themselves, but also others. Many nations have put in place policies to curb alcohol consumption. In the United States of America, the policy on drinking is a minimum legal drinking age (MLDA). The consequences of drinking underage in many states is a fine of $200 or more and community service. However, many individuals disagree with the MLDA being at 21. The Amethyst Initiative is a coalition of college chancellors and presidents who believe that the MLDA should be at 18. This paper looks at the effect of the current MLDA.

This study consists of data from the National Health Interview Sample Adult Files 1997-2007 (NHIS), where traits of individuals were gathered to monitor health of US citizens. The other data set is death certificates from the National Vital Statistics (NVS). The variables from the NHIS come from an interview where individuals report their traits. This is problematic because of the fact that individuals below the age cutoff are not likely to state their alcoholic behavior since it is illegal and can result in severe consequences. The interview asked the question if the individual drinks alcohol and not the amount, so recall bias has a minimum effect on the data.

To explain this difference, we use the regression discontinuity design. The regression discontinuity design consists of a regression with a cutoff at the MLDA to separate individuals into a control and treatment groups. The treatment is allowing individuals to drink. The treatment group is individuals above 21 and the control group is individuals below 21. Running the regressions of age on alcohol consumption and mortality gets estimates on both of them. However, the these are not the effects in interest. The effect we really want is to see if drinking alcohol leads to an increase in mortality. However, we cannot run a regression of the effect of drinking on mortality due to the fact that individuals who drink are systematically different than individuals who don’t. To estimate an effect of drinks on mortality, we use the instrumental variable method.

In the regression discontinuity design, figures are key to visualize the effect of the regression. Figures allow us to clearly see the effect of the policy and the change of what happens at the cutoff. Regressions allow us to get a quantifiable estimate of the effect of age on mortality, alcohol consumption, and motor vehicle accident. The effect of turning 21 makes people 8.7% more likely to consume alcohol. Also, turning 21 increases mortality by 9.548 deaths per 100,000 people. Using these two estimates, we are able to derive the instrumental variable estimate, the effect of drinks on mortality, which is that drinking increases death by 109.747 deaths per 100,000 people. This estimate is suspiciously big, which leads me to believe that one of the key assumptions is not met.

Data

The data set for alcohol consumption was taken from National Health Interview Sample Adult Files 1997-2007, and the data set for mortality was taken from death certificates from the National Vital Statistics. In the NHIS design, researchers went to individuals' houses and asked them to take part in the interview in order to get their information. Interviewers were only able to gather the data for people currently at their house and above 17. The questions we care about are traits of individuals and alcohol consumption of individuals. The NVS data states when the individual died and the way that the individual died. Both of these data sets are brought together to examine the effect of drinking on mortality. Furthermore, when discussing alcohol consumption, the measure used is a binary variable if the individual drinks or not. Mortality is measured in mortality rate, where we look at the amount of people that die per 100,000 people.

In the interview process, interviewers asked individuals that were from 17 to 21 if they consume alcohol. Knowing that there are severe consequences to underage drinking, individuals under 21 are likely to falsely report that they do not consume alcohol, when in reality they do. This is called desirability bias, where individuals state the socially desirable outcome, which here is people who are under 21 report not drinking. This is likely to bias the treatment effect of turning 21 on alcohol consumption upward. Due to the fact that interviewers asked the individual if they have consumed alcohol before, recall bias is not prevalent due to the fact that people will remember consuming alcohol or not. Another problem in data is the fact that age is not correctly reported. Age was recorded in months, meaning people that are right above 21 could still be considered under 21 and those under 21 can be considered above 21. This biases our estimate downwards. These two effects do not cancel each other out because it is unclear what the magnitude of each effect is.

Methods

The main method in analyzing the effect of alcohol on mortality is the regression discontinuity method. This method relies on the ability to clearly demonstrate the jump at the cutoff. To do this, we choose effect binwidth, bandwidth, and range. I chose the binwidth of 76 days because it allowed me to keep a good number of the data points without losing precision. Also, the bandwidth of age 19 to 23 allowed me to see the effect of the MDLA throughout many years and at age 21. The range effectively does the same thing, but with the outcome. Each of the graphs demonstrates the jump at the cutoff without having to sacrifice observation points.

One of the effects we are interested in is the effect of age on drinking behavior. Our hypothesis is that there is an increase in alcohol consumption when people turn 21. To test this hypothesis, we run a regression on the effect of the MDLA on alcohol consumption.

Equation 1: $Drinks_i = B_0 + B_1 Above21_i + B_2 AgeCenter + B_3 AgeCenter_i * Above21_i + E_i$

Drinks is the outcome variable that answers if the person drinks or not. *Above21* is a binary variable that indicated is the individual is above 21 or not. *AgeCenter* is the age of the person centered to 21. This regression gives us the effect of turning 21 on an individual's alcohol consumption. However, we wish to add covariates into this equation because we wish to improve the precision of the estimate. Adding covariates to the regression improves precision because it is possible that these variables can related to the outcome.

The next effect we are interested in is the effect of age on various forms of death. Due to the fact that the significant change at age 21 is the ability to drink alcohol, we believe that age can tell us the effect of what drinking has on mortality. The first regression we run is to see the effect of age only on various forms of death.

Equation 3: $Y = B_0 + B_1 Above21_i + B_2 AgeCenter + B_3 AgeCenter_i * Above21_i + E_i$

In this regression, *Y* is a place holder for the form of death we can choose. In this study, the various forms of death are *MVA (Motor Vehicle Accident), homicide, suicide, drugs, and external (external causes of death)*. Substituting any of these variables will give us the estimate of the respective form of death. The same problem is present in this regression, where there might be other variables associated with outcome variable. Adding covariates will take care of this problem.

When trying to see the effect of alcohol consumption on mortality, it is not valid to simply regress mortality on drinks. There is fundamentally something different about people who drink alcohol than those that do not. There are many theories, but one is that people drink alcohol to have a good time and not a safe time. Also, it is unethical, for obvious reasons, to design an experiment where we give people alcohol and expect them to die. So, we do the next best thing, which is the instrumental variable estimate. The instrumental variable estimate consists of the reduced form and the first stage. The reduced form estimate is the B_1 in equation 3 when the outcome is mortality and the first stage estimate is the B_1 estimate in equation 1. By dividing the reduced form by the first stage, we get the instrumental variable estimate. To get the standard error for the instrumental variable estimate, we use the delta method. This lets us know if the estimate is significant or not.

Results

Now that we have the regression equations, we wish to visualize and quantify the results. We visualize and quantify the results through the regression discontinuity design. The regression discontinuity design utilizes figures to see the effect in a clear way. These figures are chosen bandwidths and binwidths to facilitate this.

It is important to balance the traits of individuals who were interviewed to ensure that there is no significant different between the control and treatment group. We run regressions to see the effect of turning 21 on these traits. If the estimates of the covariates are statistically significant, it shows that the control and treatment group are not the same. Also, this means that there will be ways other than wage that can affect alcohol consumption. To ensure there are no other ways that can affect alcohol consumption or mortality, we balance them.

Table 1: Effect of Turning 21 on Different Traits

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	uninsured	hs diploma	hispanic	white	black	employed	married	working	going school	male
Above21	-0.017	0.014	-0.010	0.017	-0.015	0.008	-0.031***	0.011	0.008	0.017
	(0.013)	(0.011)	(0.012)	(0.015)	(0.010)	(0.014)	(0.010)	(0.014)	(0.011)	(0.015)
AgeCenter	0.026***	0.019**	0.002	-0.009	0.008	0.059***	0.052***	0.057***	-0.059***	-0.024***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.007)	(0.009)	(0.006)	(0.009)	(0.008)	(0.009)
AgeCenter Above21	-0.018	-0.017*	-0.004	0.006	-0.002	-0.027**	0.002	-0.025**	0.020**	0.028**
	(0.012)	(0.010)	(0.011)	(0.013)	(0.009)	(0.012)	(0.009)	(0.012)	(0.010)	(0.013)
Constant	0.318***	0.819***	0.243***	0.553***	0.157***	0.642***	0.153***	0.642***	0.165***	0.427***
	(0.010)	(0.008)	(0.009)	(0.010)	(0.008)	(0.010)	(0.007)	(0.010)	(0.008)	(0.010)
Observations	18,667	18,667	18,667	18,667	18,667	18,667	18,667	18,667	18,667	18,667
R-squared	0.001	0.002	0.000	0.000	0.000	0.013	0.019	0.014	0.019	0.000

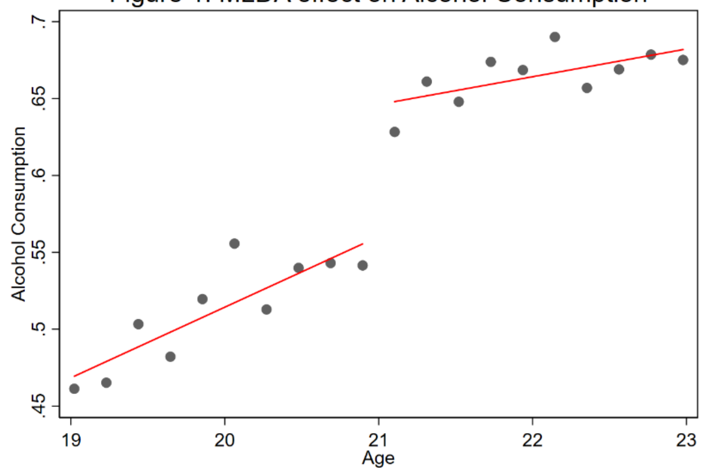
Robust standard errors in parentheses

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

In this regression table, all the variables are not significant, meaning that they do not have an impact when turning 21. This is not true for the variable *married*, which seems to have significance at the 99% level. When we run 10 regression it is likely that one of them is significant due to random chance. To fix this, we can use the Bonferroni correction, which simply states that we hold ourselves accountable to a higher standard by dividing the p-value by the number of regressions run. This is not necessary in this case as we can assume that all variables are not statistically significant and do not have an effect when people turn 21.

Visualizing both of the effects helps us to understand what happens at the cutoff. When running the regression of equation 1, we get a graph that looks like this.

Figure 1: MLDA effect on Alcohol Consumption

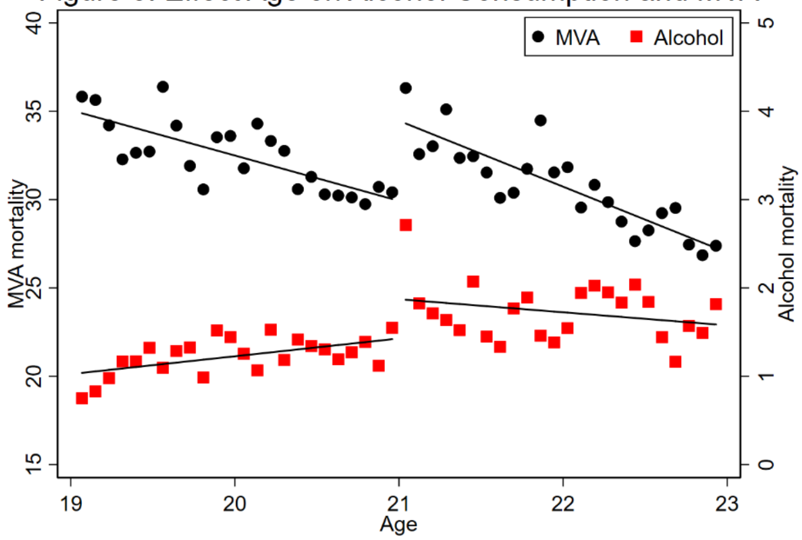


The bandwidth, binwidth, and ranges were chosen to demonstrates the jump at age 21 in alcohol consumption clearly without sacrificing many observation points. The jump shows that turning 21 increases an individual's likeliness to drink. Moreover, we wish to see the effect that turning 21 has on mortality so that we can estimate the effect of drinking on mortality. This graph is the result of running equation 3.

Figure 2: Effect of Age on Mortality

Due to the fact that the plots were scattered, the best plot for this graph would be a quadratic in order to grab the effect of data plots that are further. In this graph, there is also a jump, which demonstrates an increase in people dying when they turn 21. Another graph of interest is to see if turning 21 has an effect on motor vehicle accident (MVA). We want to know this because drunk driving is thought to be a leading cause of motor vehicle fatalities in the USA.

Figure 3: Effect Age on Alcohol Consumption and MVA



The regression in red is from figure 1; we use it again to see the effect of turning 21 on alcohol consumption and MVA side by side. Again, there is an increase in motor vehicle accidents when turning 21. Looking at these figures, it is possible to assume that turning 21 increases alcohol consumption, mortality, and MVA.

Visualization of the graphs is not enough, to get the actual effect, we need the tables that put the “jumps” into numbers. Figure 1 shows the effect of age on alcohol, and table 2 transcribes this effect into numbers.

Table 2: Effect of Turning 21 on Alcohol Consumption

VARIABLES	(1) Alcohol Consumer	(3) Alcohol Consumer	(4) Alcohol Consumer	(5) Alcohol Consumer	(6) Alcohol Consumer	(7) Alcohol Consumer
Above21	0.087***	0.090***	0.080***	0.087***	0.090***	0.082***
	(0.014)	(0.022)	(0.029)	(0.014)	(0.021)	(0.029)
AgeCenter	0.043***	-0.022	-0.055	0.043***	-0.022	-0.055
	(0.009)	(0.036)	(0.091)	(0.009)	(0.036)	(0.091)
AgeCenter Above21	-0.023*	0.097*	0.219*	-0.023*	0.095*	0.211*
	(0.012)	(0.050)	(0.126)	(0.012)	(0.050)	(0.124)
AgeCenter^2		-0.033*	-0.074		-0.033*	-0.074
		(0.018)	(0.105)		(0.018)	(0.105)
AgeCenter^2 Above21		0.006	-0.063		0.006	-0.055
		(0.024)	(0.145)		(0.024)	(0.144)
AgeCenter^3			-0.014			-0.014
			(0.035)			(0.035)
AgeCenter^3 Above 21			0.050			0.048
			(0.048)			(0.047)
birthday	0.002	0.021	0.036			
	(0.084)	(0.085)	(0.086)			
Constant	0.558***	0.537***	0.531***	0.558***	0.537***	0.531***
	(0.010)	(0.015)	(0.021)	(0.010)	(0.015)	(0.021)
Observations	18,667	18,667	18,667	18,667	18,667	18,667
R-squared	0.024	0.025	0.025	0.024	0.025	0.025

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

The effect demonstrated in figure one is the coefficent on post without the birthday effect. The birthday effect takes care of risky behavior on the person's birthday, where individuals are more likely to partake in unsafe amounts of alcohol than any other day. The estimate of the jump is the coefficent on *Above21*, which is .087. This means that turning 21 is likely to increase alcohol consumption by 8.7%. The effect of turning 21 on alcohol consumption is the same with or without the birthday effect. The birthday effect doesn't have an effect on our estimates. Figure 2 demonstrates the effect of age on mortality. Transcribing the graph into a regression table allows us to see the numerical effect of age on mortality.

Table 3: Effect of Age on Mortality

VARIABLES	(1) all	(2) mva	(3) alcohol	(4) homicide	(5) drugs	(6) external	(7) suicide
Above21	9.548***	4.663***	0.745***	0.200	0.306	8.475***	1.814**
	(1.985)	(1.155)	(0.221)	(0.522)	(0.289)	(1.922)	(0.699)
AgeCenter	-0.831	-2.933	-0.325	1.322	0.390	-2.330	0.138
	(3.290)	(1.914)	(0.366)	(0.865)	(0.479)	(3.184)	(1.158)
AgeCenter Above21	-6.017	-0.823	-0.211	-2.491**	-0.341	-4.147	-0.700
	(4.653)	(2.706)	(0.518)	(1.224)	(0.677)	(4.504)	(1.638)
AgeCenter^2	-0.840	-0.185	-0.267	0.267	0.029	-0.780	0.056
	(1.615)	(0.940)	(0.180)	(0.425)	(0.235)	(1.563)	(0.569)
AgeCenter^2 Above21	2.904	0.198	0.464*	0.148	0.138	1.855	0.031
	(2.284)	(1.329)	(0.254)	(0.601)	(0.332)	(2.211)	(0.804)
Constant	93.073***	29.809***	1.255***	17.598***	3.917***	73.005***	11.698***
	(1.404)	(0.817)	(0.156)	(0.369)	(0.204)	(1.359)	(0.494)
Observations	48	48	48	48	48	48	48
R-squared	0.682	0.722	0.574	0.394	0.723	0.510	0.489

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

In this table, we see that the effect of turning 21 on total causes of mortality (*all*) and others causes of death that are interesting. The coefficent on *Above21* is the effect of turning 21 on mortality, which shows that turning 21 will increase mortality by 9.548 deaths for every 100,000 individuals.

These estimates still do not give us the effect we want, which is the effect of drinking on mortality. Because we cannot see the regress mortality of alcohol consumption, we use the instrumental variable of age to assign people into treatment and control groups. Also, people who drink alcohol willingly are different from people who do not, so they do not provide good comparison groups. Now that we have the estimates for the reduced form and the first stage, we can divide the reduced form by the first stage to get an effect of alcohol consumption on mortality. The instrumental variable estimate is 109.747, which translate to on average people who consume alcohol die 109.747 per 100,000. Using the delta method, the standard error for the instrumental variable estimate is 29.656, demonstrating that the estimate is significant. The estimate does not make logical sense, the small population of people who comply with the study, people who only change drinking behavior at 21, are dying.

Conclusion

Due to many people believing that society can benefit if the MLDA is lowered, we look at the effect of the current MLDA though regression discontinuity design. The data from National Health Interview Sample Adult Files 1997-2007 shows data on people's alcohol behavior and shows mortality rates. However, there is bias due to the fact that the interviews were not done at random, that people who were present in their home were only ones being interviews. Moreover, we cannot put people in the treatment group and force them to drink to find the effect of alcohol. Due to this moral hazard, the variable age is used to make the study randomized. People cannot select themselves into an age category; they are either above or below the cutoff. The treatment group is the people above 21 and the control is people below 21.

To analyze the effect of age on alcohol consumption and mortality, we use the regression discontinuity design. The regression discontinuity analyzes the effect of the MDLA on mortality. People are put into the treatment group or control group based on their age. A key characteristic is figures, which help us visualize the effect of age on alcohol consumption and mortality. To effectively visualize this, effective binwidth, bandwidths, and ranges were chosen. In the figures, there are jumps at age 21, which states that there is an effect of turning 21 on alcohol consumption and mortality. Through the tables, we find that turning 21 increases alcohol consumption by 8.7% and turning 21 increases mortality by 9.548 per 100,000 individuals. However, these are not the estimates that we are interested in, we are interested in the effect of drinking on the outcome of mortality.

To get the desired effect, we use the instrumental variable method approach, which gives us an estimate of 109.747 people per 100,000. This means that out of 100,000 people, 109.747 people are going to die due to alcohol consumption. Using the delta method, the standard error for this estimate is 29.656, demonstrating that this estimate has significance. Also, this effect is only true for people who start drinking when they turn 21, people who comply to the MDLA. This estimate seems suspiciously high; therefore, we should see if the three assumptions are met. The first assumption is that the estimate of the first stage and reduced effect are not biased. This is false due to the fact that individuals under 21 falsely reported that they are not drinkers due to the consequences of underaged drinking, this biases our first stage upward. Also, age was calculated as months in NVS, leading to people right above or below the cutoff being counted as the group they are not supposed to be in, this biases the reduced form downwards. Both the reduced from and first stage are biased, so this assumption does not hold. The second assumption is that the first stage is not zero. The first stage estimate is the treatment effect of turning 21 on alcohol. This is not 0, therefore this assumption holds out. The third assumption is that the only way that turning 21 effects mortality is through alcohol consumption. In this case, alcohol consumption is not the only channel that affects mortality at 21. In this design, we assume that the only thing that changes is drinking, that people only drink when they turn 21. However, we don’t take into consideration drinking behavior, which is when, where, and how people drink. Since two of the three assumptions are not met, the instrumental variable estimate is highly inflated.

An important point to consider is that there is an increase in deaths when turning 21. The Amethyst Initiative state that reducing the drinking age will decrease mortality. Due to the findings of this study, I would disagree and state that this would only remove the effect to individuals that are18 instead of 21. Moreover, the number of individuals 18 is higher than people 21 and we do not know the risk behavior of people who are 18, so the effect on 18-year-olds can actually be greater.