**LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE** 

**Department of Psychological and Behavioural Science**

**COURSEWORK SUBMISSION FORM AND**

**PLAGIARISM/ACADEMIC HONESTY DECLARATION**

Please ensure that a completed copy of this form is uploaded as part of your coursework submission.

**Candidate Number: 46341**

**Course code: PB4A7**

**Word-count: 2720**

**Date**: 23 January 2023

The Department wishes to draw your attention to the School Calendar Regulations on Assessment Offences and Plagiarism:

<https://info.lse.ac.uk/Staff/Divisions/Academic-Registrars-Division/Teaching-Quality-Assurance-and-Review-Office/Assets/Documents/Calendar/RegulationsAssessmentOffences-Plagiarism.pdf>

All work submitted as part of the requirements for any assessment of the School (e.g., examinations, essays, dissertations, and any other work, including computer programs), whether submitted for formative or summative assessment, must be expressed in your own words and incorporate your own ideas and judgments. Plagiarism must be avoided in all such work. Plagiarism can involve the presentation of another person’s thoughts or words as if they were your own. However, please note that plagiarism also includes self-plagiarism, which is where you as the author re-use your own previously submitted work or data in a “new” written piece of work without letting the reader know that this material has appeared elsewhere. The definition of “your own work” also includes work produced by collaboration or group-work expressly permitted by the Department.

Please also note that plagiarism as defined by the School above, need not be deliberate or intentional for it to constitute an assessment offence.

**Declaration (without signature, to preserve anonymity): Having read and understood LSE’s guidelines on plagiarism/academic honesty, I hereby confirm by completing and attaching this form that the work submitted is my own. By submitting this form I hereby confirm I understand the Department’s policy on summative assessment, and that I have read the relevant parts of the MSc programme handbook.**

Table of Contents

[1. Replication 3](#_Toc125406364)

[1.1. Summary and Overview: 3](#_Toc125406365)

[1.2. Checking for Heaping/manipulation 3](#_Toc125406366)

[1.3. Checking for Covariance 4](#_Toc125406367)

[1.4. Main results – Punishment and Recidivism 5](#_Toc125406368)

[1.5. Robustness Testing 6](#_Toc125406369)

[1.5.1 Donut holes with Main Results 6](#_Toc125406370)

[1.5.2. Donut Holes with Local Polynomials 7](#_Toc125406371)

[1.5.3. Cmograms and Rdplots Excluding Outliers 8](#_Toc125406372)

[2. Discussion 10](#_Toc125406373)

[2.1 Methodological Issues 10](#_Toc125406374)

[2.1.1 Bandwidth Selection 10](#_Toc125406375)

[2.1.2. Missing Income Variable 10](#_Toc125406376)

[2.1.3. Effect of Culpability at 0.08 BAC Threshold 11](#_Toc125406377)

[2.1.4. Granularity & Applicability of Results 11](#_Toc125406378)

[2.2. Theoretical Issues 12](#_Toc125406379)

[2.2.1. Welfare Maximisation and Punishment 13](#_Toc125406380)

[2.2.2. Deterring Crime 13](#_Toc125406381)

[2.3. Extensions 13](#_Toc125406382)

[Works Cited 15](#_Toc125406383)

[Appendix 16](#_Toc125406384)

# Replication

## 1.1. Summary and Overview:

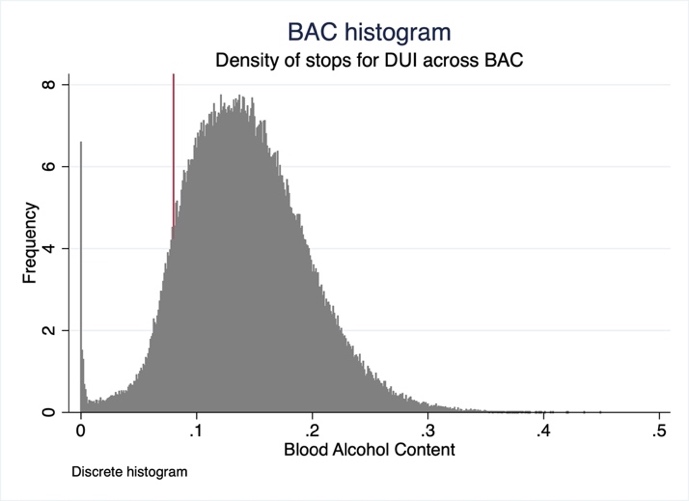
Hansen investigates the relationship between punishment and recidivism in this paper. Using data from Washington state administrative records from the years 1995-2011, he compares driving under the influence charges (DUI’s) with rates of recidivism in drunk drivers. He does this by running a regression discontinuity design (RDD), using the blood alcohol level 0.08 as the cutoff for a DUI, and 0.15 for an aggravated DUI. Using these thresholds, he investigates rates of recidivism at the DUI and aggravated DUI threshold to find the effect of punishment on recidivism. In consideration of Becker (1968)’s theory of criminal behavior, Hansen hypothesizes that recidivism will decrease as a result of punishment at the DUI threshold.

## 1.2. Checking for Heaping/manipulation

Manipulation in the data can be checked by making a continuous (Figure 1) and discrete histogram (Figure 2) of the frequency of bac1:

Figure 1 Figure 2

Chart, histogram

Description automatically generated

Note: This graph shows a discrete histogram for bac1 variable

Note: This graph shows a continuous histogram for bac1 variable

Looking at the raw data, we find that the heaping in figure 1 is a result of data that is to a degree of accuracy higher than what the breathalyzers can provide, as well as the different binning functions in the histograms. To elaborate, seefigures 3 and 4. In the data for this smaller range there is one anomaly value (.15899999). This is rounded up and included in the bar representing 0.159 in the discrete histogram (figure 3), whereas in the continuous histogram (figure 4), it is included in the bar representing 0.158 – 0.159, giving the effect of heaping in the continuous histogram. Given BAC is a discrete variable due to the limited level of accuracy of breathalyzers, we should focus on the discrete histogram, which shows no heaping.

Figure 3 Figure 4

Chart, bar chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

Note: This graph shows a discrete histogram for bac1 variable in range 0.152 – 0.161

Note: This graph shows a continuous histogram for bac1 variable in range 0.152 – 0.161

To further prove there is no manipulation, a rddensity test is run (see figure 5). Given the significance value of this is above 0.05, as can be shown in the confidence interval overlap, we can conclude there is no manipulation.

Figure 5

Chart, histogram

Description automatically generated

Note: This graph shows a rddensity graph to check for manipulation of bac1 in the dataset.

## 1.3. Checking for Covariance

RDD relies on the continuity assumption, which states that if no treatment were applied, the lines would continue smoothly past the cutoff. For this to hold, the other variables must not be covariates with DUI. To check for this, I run placebo tests for each of the control terms. As shown in Appendix table 1, accident and age are significant covariates before centring the running variable. However, they stop being significant covariates when centring the running variable, as shown in table 1. Since centring the running variable is a customary practice in RDD (Cunningham, 2021), I use a centered bac variable in any regressions.

Table 1 – Covariance Checking with Control Variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | white | male | acc | aged |
| DUI | 0.00570 | 0.00618 | -0.00335 | -0.140 |
|  | (1.14) | (1.08) | (-0.82) | (-0.85) |
|  |  |  |  |  |
| \_cons | 0.846\*\*\* | 0.784\*\*\* | 0.0834\*\*\* | 33.92\*\*\* |
|  | (206.93) | (169.39) | (25.31) | (251.95) |
| *N* | 89967 | 89967 | 89967 | 89967 |

Notes: This table contains regression discontinuity estimates of the effect of getting a DUI on characteristics of participants to check for covariance. All regressions use a bandwidth of 0.05.

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

When doing this, we see that none of the predetermined characteristics are covariates, as the P-values for the placebo test are all above the significance level (0.05). Therefore, the continuity assumption holds.

## 1.4. Main results – Punishment and Recidivism

Using a bandwidth of 0.05 and a linear model there is a significant treatment effect of dui on recidivism of -0.0240 (see table 2 and figure 6). There is also a significant quadratic relationship found (see figure 7). However, since the R2 value barely changes between these two models, the linear model seems to represent the data best. This regression was also run with smaller bandwidths of 0.025 (see panel B), in which the linear regression remained statistically significant, whilst the quadratic was not. The significance of the linear relationship across the different bandwidths suggests its robustness and adequacy to model this relationship.

*Table 2 – Regression Discontinuity Estimates for the Effect of Exceeding the 0.08 BAC Threshold on Recidivism*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Linear Regression | | | Quadratic Regression |
|  | (1) | | | (2) |
| *Panel A. BAC ∈* [0.03, 0.13] | |  |  | |
| *DUI* | -0.0240\*\*\* | | -0.0143\* | |
|  | (-5.52) | | (-2.30) | |
|  |  | |  | |
| constant | 0.106\*\*\* | | 0.100\*\*\* | |
|  | (19.78) | | | (15.97) |
| *Observations* | 89,967 | | | 89,967 |
|  |  | | |  |
| *Panel B. BAC ∈* [0.055, 0.105] |  | | |  |
| *DUI* | -0.0206\*\*\* | | | -0.0141 |
|  | (-3.58) | | | (-1.67) |
|  |  | | |  |
| constant | 0.0976\*\*\* | | | -0.943\*\*\* |
|  | (14.02) | | | (11.73) |
| *Observations* | 46,957 | | | 46,957 |
|  |  | | |  |

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Note: This table contains regression discontinuity estimates of the effect of getting a DUI on recidivism. Panel A uses a bandwidth of 0.05 and panel b uses a bandwidth of 0.25. See Appendix table 2 for results using data-driven bandwidths.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated Figure 6 Figure 7

Note: This figure shows a binned scatterplot between the mean of recidivism and bac1 with a quadratic line of best fit and bandwidths of 0.05 to either side of the cutoff.

Note: This figure shows a binned scatterplot between the mean of recidivism and bac1 with a linear line of best fit and bandwidths of 0.05 to either side of the cutoff

## 1.5. Robustness Testing

### 1.5.1 Donut holes with Main Results

A donut hole regression drops all data inside a certain bandwidth from the threshold. This removes any data that may have been manipulated, and when fitting a local polynomial creates an estimation of what the treatment effect in this threshold would have been had there been no manipulation (Cattaneo & Titiunik, 2022). This estimate can be compared to that without the donut hole. If the point estimates are similar, there is further proof that there is no manipulation.

Running a linear regression with this donut hole, one gets a significant treatment effect of -0.026 (see table 3). Comparing this with the linear regressions without the donut, we can see that the results are almost the same for both the 0.05 and 0.025 bandwidth, offering us further proof of the robustness of the data.

*Table 3 – Donut-hole Robustness Check with Main Regression*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Linear Regression | | | Linear regression with donut |
|  | (1) | | | (2) |
| *Panel A. BAC ∈* [0.03, 0.13] | |  |  | |
| *DUI* | -0.0240\*\*\* | | -0.0257\*\*\* | |
|  | (-5.52) | | (-5.36) | |
|  |  | |  | |
| constant | 0.106\*\*\* | | 0.109\*\*\* | |
|  | (19.78) | | | (15.97) |
| *Observations* | 89,967 | | | 88,085 |
|  |  | | |  |
| *Panel B. BAC ∈* [0.055, 0.105] |  | | |  |
| *DUI* | -0.0206\*\*\* | | | -0.0225\*\*\* |
|  | (-3.58) | | | (-3.40) |
|  |  | | |  |
| constant | 0.0976\*\*\* | | | .1018841\*\*\* |
|  | (14.02) | | | (13.06) |
| *Observations* | 46,957 | | | 45,075 |

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Note: This table shows the relationship between recidivism and DUIs with large bandwidths of 0.05, and smaller bandwidths of 0.025 with, and without, a donut hole kernel.

### 1.5.2. Donut Holes with Local Polynomials

Local polynomials run a nonparametric regression within a bandwidth rather than the entire dataset, meaning they have no underlying assumptions of the relationship between the variables. As a result, they are less likely to exhibit overfitting and erratic behavior near the cutoff (Cattaneo & Titiunik, 2022; Fan & Gijbels, 2018). This is crucial in a donut hole regression since we are focusing on the area near the cutoff. It is important to use high-order local polynomials when using a donut hole to pick up potential trends caused by manipulation near the threshold. By then comparing the high-order local polynomials with and without the donut hole, one can further check if there was manipulation in the dataset. I employ first, second, and third order polynomials using data-driven bandwidths with all three kernels. I found there to be no evidence of manipulation given coefficients with, and without the donut hole are almost identical (see table 4). Moreover, the similarity of the coefficients between those with data-driven bandwidths and high-order polynomials compared to the original linear regression with predetermined bandwidths further reinforce the robustness of the main regression (table 3). However, since high order local polynomials show the effect of a DUI to be below 0.02, it may hint at the original estimate of 0.02 being too high.

Table 4 – Donut Hole Robustness Check with Local polynomials

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Triangular Kernel  (1) | | Uniform Kernel  (2) | | Epanechnikov Kernel  (3) | |
| *Donut* |  | Yes |  | Yes |  | Yes |
|  |  |  |  |  |  |  |
| *Panel A:* First order  Polynomial |  |  |  |  |  |  |
| RD\_Estimate | -0.0182\*\* | -0.0196\* | -0.0179\* | -0.0187\* | -0.0183\*\* | -0.0193\* |
|  | (-3.00) | (-2.29) | (-2.58) | (-2.06) | (-3.02) | (-2.35) |
|  |  |  |  |  |  |  |
| *Observations* | 214558 | 212676 | 214558 | 212676 | 214558 | 212676 |
|  |  |  |  |  |  |  |
| *Panel B:* Second order  Polynomials |  |  |  |  |  |  |
| RD\_Estimate | -0.0175\* | -0.0203 | -0.0150 | -0.0160 | -0.0170\* | -0.0195 |
|  | (-2.24) | (-1.61) | (-1.78) | (-1.42) | (-2.13) | (-1.58) |
|  |  |  |  |  |  |  |
| *Observations* | 214558 | 212676 | 214558 | 212676 | 214558 | 212676 |
|  |  |  |  |  |  |  |
| *Panel C:* Third order polynomials |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| RD\_Estimate | -0.0168 | -0.0201 | -0.0154 | -0.0186 | -0.0174 | -0.0189 |
|  | (-1.75) | (-1.36) | (-1.50) | (-1.32) | (-1.84) | (-1.32) |
|  |  |  |  |  |  |  |
| *Observations* | 214558 | 212676 | 214558 | 212676 | 214558 | 212676 |

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Note: This table contains regression discontinuity estimates for the impact of DUIs on recidivism for different kernels and order polynomials with, and without a donut hole.

### 

### 1.5.3. Cmograms and Rdplots Excluding Outliers

Looking at figure 6, one can see there are significant outliers in values where BAC is lower than 0.06 and higher than 0.11. Using bandwidths which exclude these outliers can mitigate their effect on the overall trend found and allow for a further robustness check. This is shown in figure 9. Comparing this with the previous results found in figure 6, it becomes noticeable that the treatment effect is not as large. This further proves that the original treatment effect of punishment on recidivism may have been overestimated. Since a rectangular kernel was imposed, this could be due to these outliers flattening the relationship (Huntington-Klein, 2021), thus overestimating the treatment effect. This is further seen in the rdplot in figure 8, as the treatment effect is even smaller when using a high order local polynomial, which can offer a more precise estimate[[1]](#footnote-1).

Chart

Description automatically generated**Chart, scatter chart

Description automatically generated**Figure 8 Figure 9

Note: This graph shows an rdplot with a fourth-order polynomial fitted for the relationship between recidivism and bac1within bandwidths of 0.06 – 0.11

Note: This graph shows a binned scatterplot of recidivism and bac1 with a linear line of best fit with bandwidths of 0.06-0.11.

# Discussion

Hansen’s study has significant methodological and theoretical limitations. These will be covered sequentially, followed by suggested extensions to his study.

## 2.1 Methodological Issues

### 2.1.1 Bandwidth Selection

Hansen uses predetermined bandwidths throughout his analysis. These are present at every regression made and are included in the main regressions, and covariate regressions as well as when checking for kernel robustness and running the donut hole robustness check. In most of these, Hansen employs bandwidths of either 0.05 or 0.025. He chooses these because they include most of the dataset without overlapping with the aggravated DUI threshold.

Since Hansen’s paper, several strategies have been developed to choose data-driven bandwidths such as rdbwselect in Stata. Using data-driven bandwidths increases the replicability of the study, decreases the probability of p-hacking, and optimizes the bias-variance tradeoff (Curini, 2020). Hansen touches on his preselection of bandwidths by demonstrating that the main regression results are significant for varying bandwidths from 0.02 to 0.068. However, though my data may be different, when I implemented a local polynomial with a uniform kernel with data-driven bandwidths, the optimal bandwidths chosen are below 0.02. This is especially troubling given Hansen only tests the effects of different bandwidths with the main regression and donut hole robustness check, not with the covariance check, kernel robustness check or even when implementing a second-order polynomial in his main regression. By doing this, he fails to check covariance, kernel robustness, or the effects of using a second-order polynomial with optimal data-driven bandwidths on his main results. This reduces the robustness of his results.

### 2.1.2. Missing Income Variable

Hansen doesn’t control for income in his study. Income could be important to control for given the effect it could have on punishment, and thus recidivism. For example, those in higher income groups may have better lawyers, which can result in less harsh sentences. Moreover, being in a higher income group could affect the impact certain punishments have, such as fines and driving license suspensions, as one may be able to afford the ignition interlock license and its maintenance fee. These benefits that come with being in a higher income group may affect the relationship between punishment and recidivism between income groups. Not controlling for this affects the value of the relationship Hansen draws, as it is non-specific to income groups. Given Hansen doesn’t record the income groups of those in the study, it is unclear whether the relationship he finds is exclusive to those in high-, middle- or low-income groups, or is a general depiction of all of them.

### 2.1.3. Effect of Culpability at 0.08 BAC Threshold

A fundamental assumption of regression discontinuity design is that there is no other treatment being applied at the cut-off point. Hansen assumes that the only treatment given at the cut-off is that of punishment, meaning that the drop in recidivism at the cut-off can be solely attributed to those being punished. However, this is not necessarily the case. At the cut-off point, one is charged with a crime. Part of being charged with a crime is punishment, yet the more fundamental part is that of being told you have done something wrong. Being told you have committed wrongdoing can increase culpability and accountability for one’s actions, therefore potentially reducing recidivism, regardless of whether one is punished. This is problematic at the 0.08 cut-off, where people to the left of the cut-off (BAC <0.08) were technically doing no wrongdoing, whereas people to the right were both accused of wrongdoing and punished. By instead assuming the difference in recidivism is entirely due to punishment, Hansen ignores the psychological effects of culpability on recidivism. This could potentially mean that the effect of punishment on recidivism is lower than what Hansen finds. This could explain why the drop in recidivism is more pronounced at the 0.08 cut-off than at the 0.15 cut-off, as in the 0.08 cut-off, only those to the right were held accountable, whereas at the 0.15 cut-off people at both sides of the cut-off were both charged with a crime and punished.

### 2.1.4. Granularity & Applicability of Results

There is a clear lack of granularity in Hansen’s report which affects the external validity of the correlation he finds. Hansen does not analyze the relationship between punishment and recidivism for different punishments, ethnicities, and sexes. Moreover, the underrepresentation of minority groups in his dataset limits the applicability of his results as they predominantly represent the relationship between punishment and recidivism for white males.

#### Punishment

Firstly, Hansen’s inability to distinguish between different punishments for a DUI and their corresponding effects on recidivism significantly affects the validity of the conclusions one can draw from his report. In Washington state, a DUI can be punished in several ways such as a fine, a jail sentence or a license suspension (Hansen, 2015). These different punishments may have varying impacts on the person being punished, and thus, different impacts on their recidivism. For example, a person given a fine for driving under the influence may be more likely to recommit than a person given a jail sentence, due to their differing impacts on well-being. By ignoring this, the strength and direction of the relationship Hansen finds is less meaningful, as it only depicts the relationship between the punishments imposed in Washington state and recidivism. Since Hansen doesn’t control for what these punishments are, the conclusions drawn from his paper cannot be generalized to a greater population, as he only demonstrates a significant negative correlation for the punishments imposed in the sample (which are not known). Therefore, though a negative correlation is proven in Washington state, Hansen does not prove that all punishments will prevent recidivism and fails to establish which do in his sample. This significantly limits the investigation's applicability towards public policy decisions outside of Washington state.

#### Race

Traces of racial discrimination still permeate American society. Most importantly to this study, judicial discrimination remains prominent, as black defendants are much more likely to receive harsher sentences than white defendants (Steffensmeier et al., 1998). Discrimination in the judicial system could result in differing relationships between punishment and recidivism depending on race. By not looking at this granularly, Hansen ignores these potential differences and instead offers a general relationship. Moreover, the relationship he finds is mostly applicable to whites, as they take up 13.84 per cent of the sample size in the data used for the replication. Though different to Hansen’s data, one could assume the underrepresentation would also be a problem in his given 78% of the population in Washington state is white (*Population by race,* 2021).

#### Gender

There is also a lack of granularity when investigating the effects of punishment on recidivism for different genders. Given females are generally given more lenient sentences in court than males (Doerner & Demuth, 2014), not looking at the effects of punishment for each gender independently may limit the applicability of the results in public policy. Moreover, as in race, there is a vast underrepresentation of females in Hansen’s sample size, thus limiting how applicable the relationship he finds is to women at all.

## 2.2. Theoretical Issues

### 2.2.1. Welfare Maximisation and Punishment

In his study, Hansen concludes that since exceeding the DUI threshold reduces recidivism by up to two percentage points, punishment is “effective” at reducing recidivism. However, there is a question of efficiency that Hansen fails to touch on. Though a significant effect is found, does a 2-percentage point decrease in recidivism truly constitute an “effective” policy? Considering the negative impacts on well-being that some of the punishments that can be given with a DUI can have (Armstrong & Weaver, 2010), this two-percentage point decrease may instead point us in the opposite direction from Hansen’s conclusion. Perhaps, such a small percentage point difference, though significant, warrants a different approach to tackling recidivism, which doesn’t result in such an extreme drop in welfare.

### 2.2.2. Deterring Crime

Yet another problem with Hansen’s study is that while he finds significant evidence suggesting punishment can deter drunk drivers from re-committing, there is no evidence regarding the effectiveness of punishments in deterring those from drunk driving in the first place. Therefore, Hansen does not give an overview of punishment and deterrence, but instead only shows convincing evidence for the effect that punishment can have on deterring people from re-committing.

## 2.3. Extensions

To build on the flaws of Hansen’s study, a similar study with higher granularity on the effects of punishment on recidivism depending on race, gender, punishment, and income group could be done. Given the demographic of Washington state is predominantly white, this study could be done in a state with a more varied demographic, such as California (Johnson et al., 2022). This would allow for a more varied range of data and the ability to granularly assess the effect of punishment on recidivism for different races, sexes, income groups, and different punishments. This would greatly increase the applicability of the results to different states and inform more specific public policy decisions. Most importantly, this would also allow policymakers to see what punishment works best. This could be investigated by comparing regression discontinuity effects of different types of punishments such as jail sentences, fines, driving license suspension or even offender rehabilitation on recidivism. This is especially important considering the recent evidence suggesting the ineffectiveness of jail sentences in reducing recidivism (Bales & Piquero, 2011).

Given the effect that punishment has on recidivism which Hansen finds is relatively small, a natural experiment to follow his investigation could assess the impact of using alternative policies to reduce recidivism such as “early interventions”. These interventions aim to deter potential crime-doers through educational campaigns or even therapy administration for troubled children. Rather than using an RDD approach, these could be investigated using a differences-in-difference approach, whereby the effects of their implementation are compared between states in which it has been implemented and those in which it hasn’t. This would also avoid the obvious downfalls of RDD.

# Works Cited

Armstrong, S., & Weaver, B. (2010). What Do the Punished Think of Punishment? The comparative experience of short prison sentences and community-based punishments.

Bales, W.D. and Piquero, A.R. (2011) “Assessing the impact of imprisonment on recidivism,” Journal of Experimental Criminology, 8(1), pp. 71–101. Available at: <https://doi.org/10.1007/s11292-011-9139-3>.

Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. Annual Review of Economics, 14, 821-851. <https://doi.org/10.1146/annurev-economics-051520-021409>

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A practical introduction to regression discontinuity designs: Foundations. Cambridge University Press. <https://doi.org/10.1017/9781108684606>

Cunningham, S. (2021). Causal inference. In Causal Inference. Yale University Press. <https://doi.org/10.12987/9780300255881>

Curini, L. (2020). The SAGE handbook of research methods in political science and international relations. Sage.

Doerner, J. K., & Demuth, S. (2014). Gender and sentencing in the federal courts: Are women treated more leniently?. Criminal Justice Policy Review, 25(2), 242-269. <https://doi.org/10.1177/0887403412466877>

Fan, J., & Gijbels, I. (2018). Local polynomial modelling and its applications. Routledge. <https://doi.org/10.1201/9780203748725>

Hansen, Benjamin. "Punishment and deterrence: Evidence from drunk driving." American Economic Review 105.4 (2015): 1581-1617.

Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. Chapman and Hall/CRC.

|  |
| --- |
|  |

Population by race. Office of Financial Management. (2021, November 12). Retrieved January 18, 2023, from <https://ofm.wa.gov/washington-data-research/statewide-data/washington-trends/population-changes/population-race>

Steffensmeier, D., Ulmer, J., & Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. Criminology, 36(4), 763-798. https://doi.org/10.1111/j.1745-9125.1998.tb01265.x

# Appendix

Appendix Table 1 – Placebo tests for covariates without centering running variable

Table 1 – Covariance checking with control variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | white | male | acc | aged |
|  |  |  |  |  |
| DUI | 0.00445 | -0.0184 | -0.154\*\*\* | -6.224\*\*\* |
|  | (0.25) | (-0.93) | (-10.11) | (-10.62) |
|  |  |  |  |  |
| \_cons | 0.840\*\*\* | 0.801\*\*\* | 0.171\*\*\* | 39.45\*\*\* |
|  | (58.94) | (50.32) | (13.58) | (81.67) |
| *N* | 89967 | 89967 | 89967 | 89967 |

Notes: This table contains regression discontinuity estimates of the effect of getting a DUI on characteristics of participants to check for covariance. All regressions use a bandwidth of 0.05.

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Appendix Table 2 – Rdrobust local polynomial regression estimates for the effect fo punishment on recidivism past the 0.08 BAC Threshold.

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
|  | Linear | Quadratic |
| RD\_Estimate | -0.0179\*\* | -0.015 |
|  | (-2.58) | (-1.78) |
| *N* | 214558 | 214558 |

*Note:* This table contains the regression results when running rdrobust local polynomials to find the relationship between punishment and recidivism. Both regressions use data-driven bandwidths to optimse bias-variance tradeoff. Based on data from the 1997-2007 Washington State Impaired Driver program. Standard errors in parantheses. (1) – Linear rdrobust regression results, (2) quadratic rdrobust regression results.

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Figure 1

Chart

Description automatically generated

Note: This graph shows an rdplot with a third-order polynomial fitted for the relationship between recidivism and bac1within bandwidths of 0.06 – 0.11

Figure 2

Figure 2

***Chart

Description automatically generated with medium confidence***

Note: This graph shows an rdplot with a second-order polynomial fitted for the relationship between recidivism and bac1within bandwidths of 0.06 – 0.11

Figure 3

***Diagram

Description automatically generated with medium confidence***

Note: This graph shows an rdplot with a first-order polynomial fitted for the relationship between recidivism and bac1within bandwidths of 0.06 – 0.11

Do file

//Note: using esttab package for output of tables

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* 1.2. Checking for heaping/manipulation (Question 1)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

gen dui = 0

//Creating dummy variable for DUI. DUI = 1 if past 0.08 threshold, 0 otherwise. Include "." for missing values

replace dui = 1 if bac1>=0.08 & bac1~=.

\* Centering running variable

gen bac1\_cntr = bac1-0.08 // centering bac1 value

gen bac1\_cntr\_sq = bac1\_cntr^2 //creating centered bac1 squared variable

histogram bac1, discrete width(0.001) ytitle(Frequency) xtitle(Blood Alcohol Content) xline(0.08) title(BAC histogram) subtitle(Density of stops for DUI across BAC) note(Discrete histogram) color(gray) graphregion(color(white)) // Making discrete histogram with a white background to check for visible heaping

histogram bac1, width(0.001) ytitle(Frequency) xtitle(Blood Alcohol Content) xline(0.08) title(BAC histogram) subtitle(Density of stops for DUI across BAC) note(Continuous histogram) color(gray) graphregion(color(white)) // Making continuous histogram as well

// In order to check for heaping, we can run two histograms, one using bac1 as a discrete variable, whereby it takes on a fixed number of potential values, and one using bac1 as a continuous variable, whereby it can take an infinite numbe of values. Using bac1 as the continuous variable, we can observe heaping shown by the spikes in the graph.

// The reason this heaping is only visible on the continuous graph is because the continuous graph assumes the data can take an infinite values. This means each bar in the histogram represents a range of possible values (in this case in increments of 0.001, as we have stated in the width).

histogram bac1 if bac1>=0.152 & bac1<=0.161, discrete width(0.001) ytitle(Frequency) xtitle(Blood Alcohol Content) xline(0.08) title(BAC histogram) subtitle(Density of stops for DUI across BAC) note(Discrete histogram and bandwidth of 0.152 - 0.161) color(gray) lcolor(white) graphregion(color(white)) //Zoomed in verison of discrete histogram

histogram bac1 if bac1>=0.152 & bac1<=0.161, width(0.001) ytitle(Frequency) xtitle(Blood Alcohol Content) xline(0.08) title(BAC histogram) subtitle(Density of stops for DUI across BAC) note(Continuous histogram and bandwidth of 0.152 - 0.161) color(gray) lcolor(white) graphregion(color(white)) // Zoomed in version of continuous histogram

rddensity bac1, c(0.08) graph\_opt(graphregion(color(white))title(Rddensity graph) subtitle("Density Discontinuity Testing") xtitle("Blood Alcohol Content") ytitle(Density) leg(off)) plot

// rddensity results in p value < 0.05 (significance level), meaning there is no manipulation near the threshold. rddensity uses data-driven bandwidths to check for heaping by comparing values of left and right of cutoff

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* 1.3. Checking for covariance / continuity assumption (Question 2)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

// RDD works under the continuity assumption (assumption that lines would continue smoothly on either side of the threshold if the cutoff didnt exist)

// In order for the continuity assumption to be true, its inmportant to check that no other variables jump at the cutoff point, as this would violate the continuity assumption

// To check for this, you can run regression with the potential covariants and the interaction term dui\*bac1. If any of these regressions have a significant P-value, the continuity assumption is being violated.

eststo clear

reg white dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05, robust //results in p value higher than significance level for T-test

eststo

reg male dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05, robust //results in p value higher than significance level for T-test

eststo

reg acc dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05, robust //results in p value higher than significance level for T-test

eststo

reg aged dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05, robust //results in p value higher than significance level for T-test

eststo

esttab using "table\_covariance\_centered.rtf"

//Given there is covariance between accidents on scene and dui, as well as age and dui, in the rdd we must control for both of these variables.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*1.4. Main results - Punishment and recividism (Question 3)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//Replicating main results:

//Using standard local polynomial with predetermined bandwidths

//Here we are running a regression with controls to check the main relationship Hansen draws between punbishment and recividism. We first check for a linear relationship using bac1 and then a quadratic regression using bac1\_sq

//Given you need to center your running variable for accurate interaction term regression results, I use the centered variable bac1\_cntr

//I run these regressions first with the larger bandwidth Hansen uses of 0.05, and then with the smaller bandwidth of 0.025

eststo clear

//running linear regression with large bandwidth

reg recidivism white male aged acc dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05, robust

eststo

//running quadratic regression with large bandwidth

reg recidivism white male aged acc dui##c.(bac1\_cntr bac1\_cntr\_sq) if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05, robust

eststo

esttab using "main\_results\_large\_bandwidth\_centered.rtf"

eststo clear

\* Slightly smaller bandwidth of 0.055 to 0.105

//running linear regression with small bandwidth

reg recidivism white male aged acc dui##c.bac1\_cntr if bac1\_cntr>=-0.025 & bac1\_cntr<=0.025, robust

eststo

//running quadratic regression with large bandwidth

reg recidivism white male aged acc dui##c.(bac1\_cntr bac1\_cntr\_sq) if bac1\_cntr>=-0.025 & bac1\_cntr<=0.025, robust

eststo

esttab using "main\_results\_small\_bandwidth\_centered.rtf"

//To depict the results, running cmograms with linear and quadratic lines of best fit

cmogram recidivism bac1 if bac1>0.03 & bac1<0.13, cut(0.08) scatter line(0.08) graphopts(bgcolor(white) title(Cmogram) subtitle(Binned Scatter Graph of Means) note(linear line of best fit)) lfitci

cmogram recidivism bac1 if bac1>0.03 & bac1<0.13, cut(0.08) scatter line(0.08) graphopts(bgcolor(white) title(Cmogram) subtitle(Binned Scatter Graph of Means) note(quadratic line of best fit)) qfitci

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* 1.5.1. Robustness testing with donut holes (Question 4)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//Question 4, Implementing donut hole with standard regresison

gen donut = 0 //creating dummy variable for donut hole kernel

replace donut = 1 if bac1\_cntr>=-0.001 & bac1\_cntr<=0.001 //making dummy variable equal 1 if values lie inside bandwidth we want to drop

eststo clear

//Large bandwidth

reg recidivism white male aged acc dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<=0.05 & donut==0, robust //conditional if statement to not include values inside donut hole

eststo

reg recidivism white male aged acc dui##c.bac1\_cntr if bac1\_cntr>=-0.05 & bac1\_cntr<0.05, robust //comparing without donut hole kernel

eststo

//Small bandwidth

reg recidivism white male aged acc dui##c.bac1\_cntr if bac1\_cntr>=-0.025 & bac1\_cntr<=0.025 & donut==0, robust //conditional if statement to not include values inside donut hole

eststo

reg recidivism white male aged acc dui##c.bac1\_cntr if bac1\_cntr>=-0.025 & bac1\_cntr<=0.025, robust //comparing without donut hole kernel

eststo

esttab using "donut\_hole\_check\_main\_regression.rtf"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* 1.5.2. Local polynomials with donut holes

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//Running donut hole robustness check with local polynomials

//Using rdrobust because of its data-driven bandwidth selection which optimises the bias-variance tradeoff. rdrobust also has further smoothing functions included in the documentation

\*\*Comparing with and without donut hole with 1st order local polynomials with data-driven bandwidths\*\*

eststo clear

//Triangular Kernel

rdrobust recidivism bac1\_cntr, kernel(triangular) masspoints(off) p(1) c(0) covs(aged white acc male) //Controls make no difference

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(triangular) masspoints(off) p(1) c(0) covs(aged white acc male)

eststo

//Uniform Kernel

rdrobust recidivism bac1\_cntr, kernel(uniform) masspoints(off) p(1) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(uniform) masspoints(off) p(1) c(0) covs(aged white acc male)

eststo

//Epanechnikov Kernel

rdrobust recidivism bac1\_cntr, kernel(epanechnikov) masspoints(off) p(1) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(epanechnikov) masspoints(off) p(1) c(0) covs(aged white acc male)

eststo

esttab using "first\_order\_kernel\_rdrobust\_centered.rtf"

eststo clear

\*\*Comparing with and without donut hole with 2nd order local polynomials with data-driven bandwidths\*\*

//Triangular Kernel

rdrobust recidivism bac1\_cntr, kernel(triangular) masspoints(off) p(2) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(triangular) masspoints(off) p(2) c(0) covs(aged white acc male)

eststo

//Uniform Kernel

rdrobust recidivism bac1\_cntr, kernel(uniform) masspoints(off) p(2) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(uniform) masspoints(off) p(2) c(0) covs(aged white acc male)

eststo

//Epanechnikov Kernel

rdrobust recidivism bac1\_cntr, kernel(epanechnikov) masspoints(off) p(2) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0 , kernel(epanechnikov) masspoints(off) p(2) c(0) covs(aged white acc male)

eststo

esttab using "second\_order\_kernel\_rdrobust\_centered.rtf"

\*\*Comparing with and without donut hole with 3rd order local polynomials with data-driven bandwidths\*\*

eststo clear

//Triangular Kernel

rdrobust recidivism bac1\_cntr, kernel(triangular) masspoints(off) p(3) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(triangular) masspoints(off) p(3) c(0) covs(aged white acc male)

eststo

//Uniform Kernel

rdrobust recidivism bac1\_cntr, kernel(uniform) masspoints(off) p(3) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0, kernel(uniform) masspoints(off) p(3) c(0) covs(aged white acc male)

eststo

//Epanechnikov Kernel

rdrobust recidivism bac1\_cntr, kernel(epanechnikov) masspoints(off) p(3) c(0) covs(aged white acc male)

eststo

rdrobust recidivism bac1\_cntr if donut == 0 , kernel(epanechnikov) masspoints(off) p(3) c(0) covs(aged white acc male)

eststo

esttab using "third\_order\_kernel\_rdrobust\_centered.rtf"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* 1.5.3. - Cmograms and Rdplots Excluding Outliers

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//Not using bac1\_cntr for visualisation purposes

rdplot recidivism bac1 if bac1>0.06 & bac1<0.11, masspoints(off) c(0.08) ci(95) shade graph\_options(title(RD Plot Recidivism and BAC)ytitle("Mean of Recidivism")xtitle("BAC1")graphregion(color(white))) //Fourth order rplot

cmogram recidivism bac1 if bac1>0.06 & bac1<0.11, cut(0.08) scatter line(0.08) graphopts(title(Cmogram) subtitle(Binned Scatter Graph of Means) note(Bandwidths 0.06 - 0.11) bgcolor(white)) lfitci //cmograms nonparametric representaiton. We decide to employ

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Appendix code

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//Appendix table 1

eststo clear

reg white dui##c.bac1 if bac1>=0.03 & bac1<=0.13, robust //results in p value higher than significance level for T-test

eststo

reg male dui##c.bac1 if bac1>=0.03 & bac1<=0.13, robust //results in p value higher than significance level for T-test

eststo

reg acc dui##c.bac1 if bac1>=0.03 & bac1<=0.13, robust //results in p value lower than significance level for T-test, means significance co-variant

eststo

reg aged dui##c.bac1 if bac1>=0.03 & bac1<=0.13, robust //results in p value lower than significance level for T-test, means significance co-variant

eststo

esttab using "table\_covariance.rtf"

//Appendix table 2

eststo clear

rdrobust recidivism bac1, kernel(uniform) masspoints(off) p(1) c(0.08) covs(white male aged acc)

eststo

rdrobust recidivism bac1, kernel(uniform) masspoints(off) p(2) c(0) covs(white male aged acc)

eststo

esttab using "main\_results\_rdrobust.rtf"

//Appendix figures 1-3

//Figure 1

rdplot recidivism bac1 if bac1>=0.06 & bac1<=0.11, p(3) masspoints(off) c(0.08) ci(95) shade graph\_options(title(RD Plot Recidivism and BAC)ytitle("Mean of Recidivism")xtitle("BAC1")subtitle(3rd order rdplot) graphregion(color(white))) //Third-order rplot (in appendix)

//Figure 2

rdplot recidivism bac1 if bac1>=0.06 & bac1<=0.11, p(2) masspoints(off) c(0.08) ci(95) shade graph\_options(title(RD Plot Recidivism and BAC)ytitle("Mean of Recidivism")xtitle("BAC1")subtitle(2nd order rdplot)graphregion(color(white))) // Second-order rdplot (in appendix))

//Figure 3

rdplot recidivism bac1 if bac1>=0.06 & bac1<=0.11, p(1) masspoints(off) c(0.08) ci(95) shade graph\_options(title(RD Plot Recidivism and BAC)ytitle("Mean of Recidivism")xtitle("BAC1")subtitle(1st order rdplot)graphregion(color(white))) // First-order rdplot (in appendix)

1. Consistent in rdplots with lower-order polynomials (see Appendix figures 1-3) [↑](#footnote-ref-1)