

Candidate number: 49045

**Replication of Hansen (2015) Punishment and Deterrence: Evidence from
Drunk Driving**

Submitted as the summative assessment for
PB4A7: Quantitative Applications for Behavioural Science 2022

In the study *Punishment and Deterrence: Evidence from Drunk Driving* (**Hansen, 2015**), Hansen investigated the effect of punishments and sanctions on reducing repeat drunk driving. The study implemented a quasi-experimental design, utilising the administrative records from 1995 to 2011 of the state of Washington, U.S., where two thresholds of blood alcohol content (BAC) are used to determine the status of driving under the influence (DUI). Specifically, a driver with a measured BAC over 0.08 is considered a case of DUI and will be punished via measures such as fines, jail time, and driving license suspension. One with a BAC over 0.15 is considered a case of aggravated DUI, to which more severe punishments and sanctions are applied. Since BAC measure has clear-cut numeric thresholds for determining whether a driver will receive harsher punishments, and neither drivers or police can manipulate this measure, Hansen applied a regression discontinuity design to analyse the data (**expand on the justification of using RDD**). He hypothesized that receiving harsher punishments and sanctions at both thresholds would reduce offenders' future recidivism of DUI. He further applied the RDD to analyse the effect of receiving harsher punishments on the degree of deterrence, incapacitation, and rehabilitation in order to identify the mechanism through which harsher punishments might reduce recidivism (**check the paper to see whether this sentence is precise**).

Ideally also the results

The present study aims to replicate the findings from **Hansen (2015)** using regression discontinuity design applied to a similar data (**what is it exactly...**). It will be focused on estimating the effect of receiving punishments on reducing recidivism at the 0.08 BAC threshold only. The paper proceeds as follows. Section 1 discusses the econometric method and assumptions underlying its application to the current data. Section 2 presents the main results, and Section 3 discusses critiques and extensions to the original study. Section 4 concludes.

1 Methods and Assumptions

1.1 Assumptions of the regression discontinuity design

The present study applies the regression discontinuity design to the data provided by the class instructor. Several assumptions need to be met so that the regression discontinuity design can give accurate estimates. First, people need to be randomly assigned to receiving punishment at the BAC threshold. In other words, neither the drivers nor the police can manipulate the BAC level and thus whether one will receive punishment. **Hansen (2015)** has already given a plausible theoretical justification for this assumption. Empirically, I plot the distribution of BAC and test for discontinuity in its distribution at the threshold. Figure 1 shows the distribution of BAC level. Based on a recently developed local polynomial density estimator (**CJM, 2020**), the hypothesis that the distribution is continuous at 0.08 cannot be rejected at the level of 0.05 ($p = 0.890$). Therefore, there is no evidence for the existence of manipulation on the BAC level.

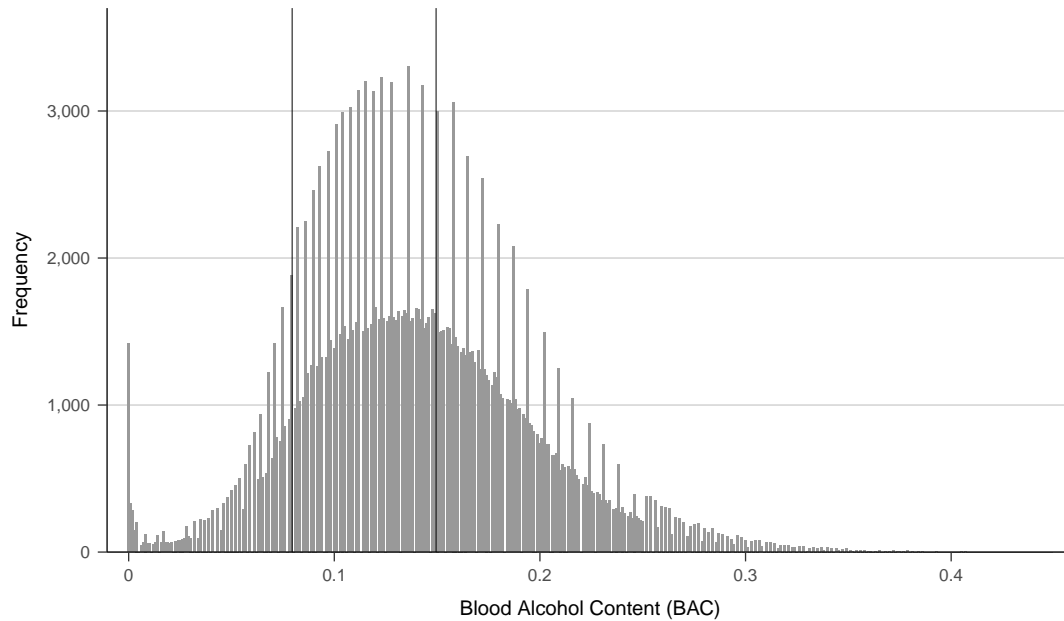


Figure 1: Histogram of blood alcohol content as a continuous variable. The blood alcohol content is plotted as a continuous variable with a bin width of 0.001, the precision used on the breathalysers. The y-axis represents the frequency of observations in each bin. The vertical black lines represent the two thresholds at 0.08 and 0.15.

The second assumption is that the running variable should not contain non-random heaping, which can lead to biased estimates in regression discontinuity models (**Barreca et al., 2011**). In other words, BAC should not be much more likely to take certain values than others. Based on Figure 1, this assumption seems to be violated. Curiously, when BAC is plotted as a discrete variable, non-random heaping disappears (Figure 2). A closer inspection of the data suggests that this is probably due to the lack of precision in BAC in the current data. Many BAC values deviate by a very small amount from the values that are supposed to be given by the breathalysers (i.e., precise to three digits after the decimal point). Some bins have a value at the left boundary deviating upwards and a value at the right boundary deviating downwards. Such bins will contain a larger number of observations than other bins and thus leads to heaping. This would not occur if the data records the values precisely so that the number of observations are rightly separated into and plotted by two bins, or if we treat BAC as discrete and give each value its own bin. **(consider using a simulation to justify this claim?)**

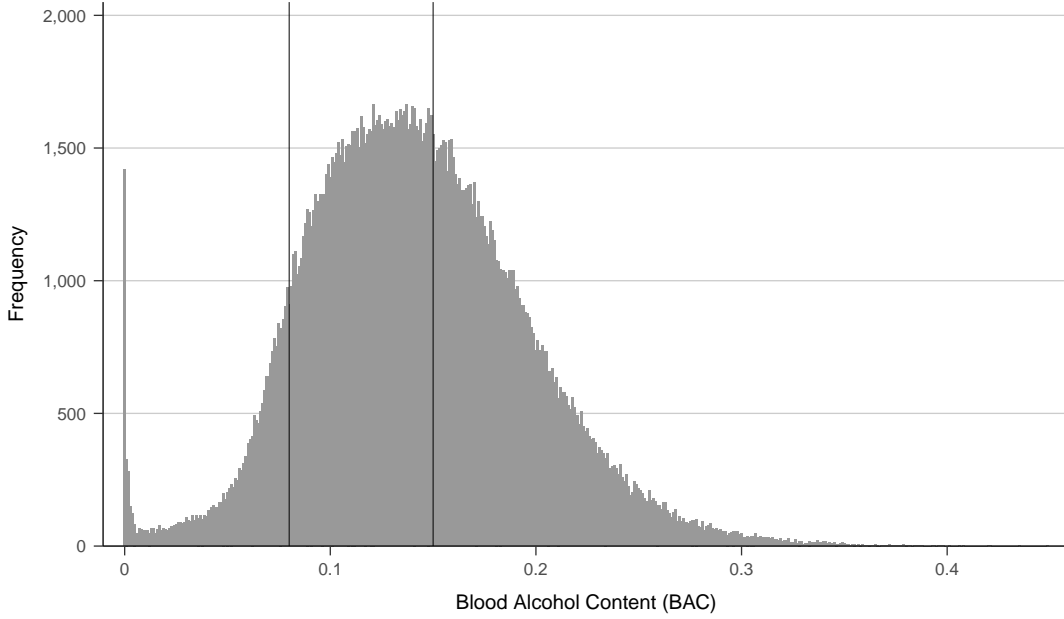


Figure 2: Histogram of blood alcohol content as a discrete variable. The blood alcohol content is plotted as a discrete variable with a bin width of 0.001, the precision used on the breathalysers. The y-axis represents the frequency of observations in each bin. The vertical black lines represent the two thresholds at 0.08 and 0.15.

1.2 The model

The present study utilises a local linear regression model with a rectangular kernel weight function to estimate the effect of receiving punishments at the threshold on recidivism. For sensitivity analyses, the models are re-estimated using second-order polynomials and triangle kernel function). The model is specified by the following function:

$$R_i = \alpha + \beta_1 DUI_i + \beta_2 BAC_i + \beta_3 DUI_i \times BAC_i + \tau Z_i + \epsilon_i \quad (1)$$

where the variable DUI_i is an indicator of whether BAC is above the 0.08 threshold, BAC_i is the measure of BAC level, R_i is an indicator of recidivism, and Z_i is a vector of control variables. The variable BAC_i in the model is centered around the threshold value of 0.08 so that the coefficients directly reflect the estimates of the effect.

1.3 Inclusion of control variables

In order to determine whether any covariates should be included in the model, I run preliminary analyses on the effect of receiving punishments at the threshold on four predetermined characteristics, including three demographic variables (gender, race, and age) and the BAC test being conducted in a traffic accident. The analyses use the same local linear regression model specified in equation (1), with a bandwidth of 0.05 and no kernel weight function. As Table 1 shows, I fail to reject any of the null hypotheses that predetermined

Table 1: Regression Discontinuity Estimates of the Effect of Receiving Punishments at the 0.08 BAC Threshold on Predetermined Characteristics

	Male	White	Age	Accident
<i>DUI</i>	0.006 (0.006)	0.006 (0.005)	-0.141 (0.164)	-0.003 (0.004)
Mean	0.784	0.846	0.085	33.99
Num. of Obs.	89,967	89,967	89,967	89,967

Note. Regression discontinuity based estimates of the effect of receiving punishments at the 0.08 BAC threshold on four predetermined characteristics. All models use a bandwidth of 0.05 and a rectangular kernel weight function. Counterfactual predictions of mean recidivism is calculated at the 0.079 BAC threshold. Heteroscedasticity-robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

characteristics remained the same at the threshold. This indicates that gender, race, age, and the BAC test being conducted in a traffic accident, on average, were not different between people who did not receive punishments and those who received punishments at the threshold¹. According to **Calonico et al. (2019)**, I will include the four predetermined characteristics in regression discontinuity models to improve the estimation precision of the effects of receiving punishments at the threshold on recidivism.

2 Results

2.1 The Effect of Punishments on Recidivism

Table 2 reports the estimated effect of receiving punishments at the threshold on recidivism. Local linear regression model with a rectangular kernel function gives the estimate that receiving punishments at the 0.08 BAC threshold decreases recidivism by 2.4 percentage points, which is statistically significant at the level of 0.01. Local second-order polynomial with a rectangular kernel function gives an estimate of a decrease in recidivism by 1.4 percentage points, which is statistically significant at the level of 0.05. These estimates are consistent across both bandwidths and across models with different types of kernel functions.

Figure 3 plots means of recidivism in bins and predicted recidivism based on linear regression models or second-order polynomials using observations within the interval $BAC \in [0.06, 0.11]$. Panel A and B and Panel C and D use different methods for binning observations and representing confidence intervals to ensure the robustness of the visual evidence (See the figure caption). All panels show an apparent drop in recidivism at the BAC threshold of 0.08. This visually indicates that there is an effect of receiving punishments at the BAC threshold in decreasing recidivism.

¹This is an additional evidence for the absence of manipulation at the threshold, which further assures the regression discontinuity design can deliver unbiased estimates.

Table 2: Regression Discontinuity Estimates of the Effect of Receiving Punishments at the 0.08 BAC Threshold on Recidivism

	Rectangular kernel		Triangular kernel	
	Linear	Quadratic	Linear	Quadratic
<i>A. $DUI \in [0.03, 0.13]$</i>				
<i>DUI</i>	−0.024*** (0.004)	−0.014** (0.006)	−0.020*** (0.005)	−0.014** (0.006)
Mean	0.104	0.099	0.100	0.099
Controls	Yes	Yes	Yes	Yes
Num. of Obs.	89,967	89,967	89,967	89,967
<i>B. $DUI \in [0.055, 0.105]$</i>				
<i>DUI</i>	−0.021*** (0.006)	−0.014* (0.008)	−0.018*** (0.006)	−0.016* (0.009)
Mean	0.101	0.098	0.101	0.100
Controls	Yes	Yes	Yes	Yes
Num. of Obs.	46,957	46,957	46,957	46,957

Note. Regression discontinuity based estimates of the effect of receiving punishments at the 0.08 BAC threshold on recidivism. Panel A presents estimates based on a 0.05 bandwidth, and Panel B presents estimates based on a 0.025 bandwidth. The table includes results from both linear and quadratic models, with either a rectangular or a triangular kernel weight function. Controls include individuals' gender, race, age, and an indicator of whether the BAC test was conducted in a traffic accident. Counterfactual predictions of mean recidivism are calculated at the 0.079 BAC threshold and mean age of the respective populations, averaging over individuals' gender and race, as well as over whether the BAC testing was conducted in an accident. Heteroscedasticity-robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

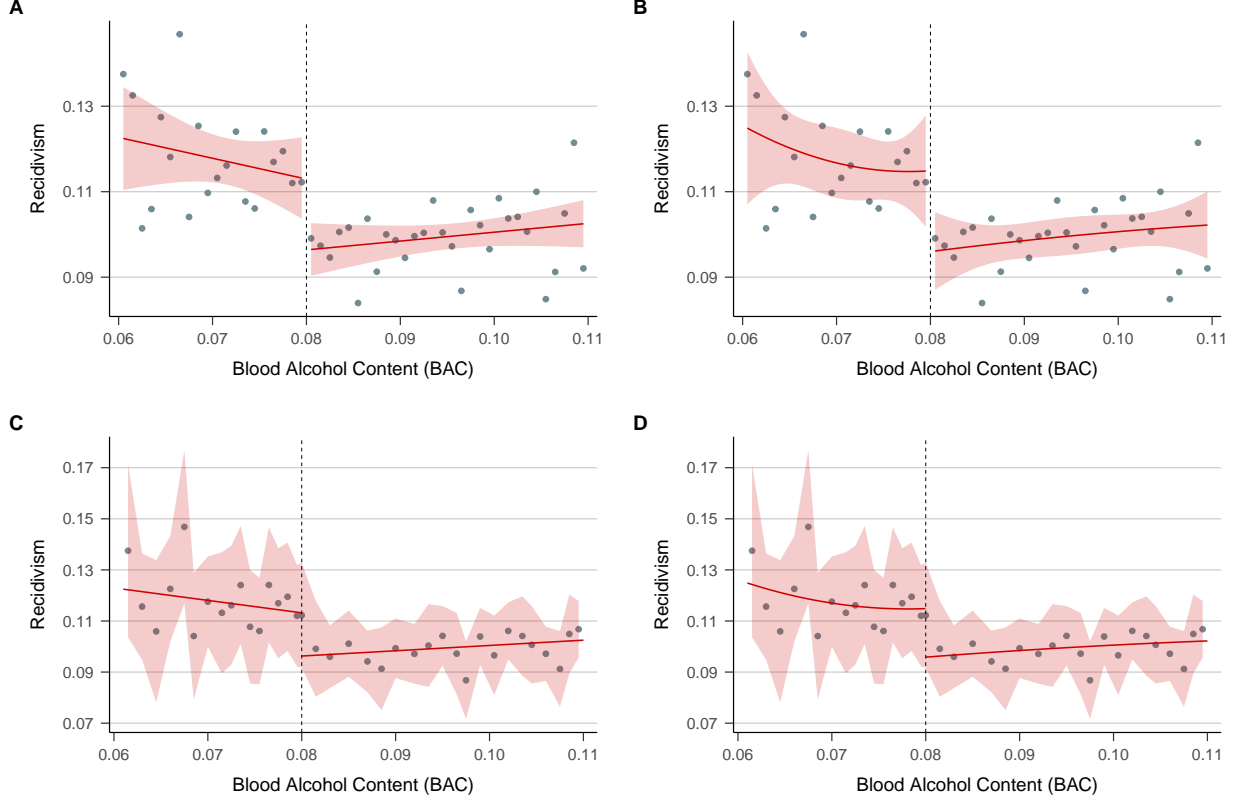


Figure 3: Regression Discontinuity Plot of the Effect of Receiving Punishments at the 0.08 BAC Threshold on Recidivism. Plot of means of recidivism in bins and predicted recidivism using simple regression models. In plotting the binned means (gray points), Panel A and B choose the number of bins on each side of the threshold using the formula $\min\{\sqrt{n}, 10 \times \frac{\ln n}{\ln 10}\}$, while Panel C and D choose the number using quantile spaced binning ***(Calonico et al., 2015)****. For the confidence intervals (shaded areas in red), Panel A and B plot the 95% confidence interval of the regression models, while Panel C and D plot the 95% confidence interval of each bin. For the regression models (red lines), Panel A and C plot the best fitted linear model, while Panel B and D plot the best fitted second-order polynomial. All panels plot the data within the interval $BAC \in [0.06, 0.11]$. The vertical dashed lines represent the BAC threshold at 0.08.

2.2 Robustness

Since there seems to be non-random heaping in the current data (Figure 1), this section will use donut regression discontinuity models to investigate the robustness of the results presented in the last section. Donut regression discontinuity models entirely drop the observations near the threshold. Under appropriate assumptions, they are effective in preventing a heap just at the threshold from biasing the estimates (**Barreca et al., 2011**).

For the present study, I drop the observations in the interval $BAC \in [0.079, 0.081]$, and re-estimate local linear and quadratic models with either a rectangular or a triangular kernel weight function. It is worth stressing that robustness test using local polynomials is especially important in a donut regression discontinuity design. This is because a donut regression model gets rid of the very observations on which the estimates of local average effects rely upon. The estimates given by donut regression models are essentially

Table 3: Donut Regression Discontinuity Estimates of the Effect of Receiving Punishments at the 0.08 BAC Threshold on Recidivism

	Rectangular kernel		Triangular kernel	
	Linear	Quadratic	Linear	Quadratic
<i>A. $DUI \in [0.03, 0.13]$</i>				
<i>DUI</i>	−0.026*** (0.005)	−0.014* (0.007)	−0.022*** (0.005)	−0.018** (0.006)
Mean	0.105	0.099	0.102	0.101
Controls	Yes	Yes	Yes	Yes
Num. of Obs.	88,085	88,085	88,085	88,085
<i>B. $DUI \in [0.055, 0.105]$</i>				
<i>DUI</i>	−0.022*** (0.007)	−0.015 (0.011)	−0.020*** (0.007)	−0.020 (0.012)
Mean	0.103	0.099	0.103	0.104
Controls	Yes	Yes	Yes	Yes
Num. of Obs.	45,075	45,075	45,075	45,075

Note. Regression discontinuity based estimates of the effect of receiving punishments at the 0.08 BAC threshold on recidivism after dropping observations in the interval $[0.079, 0.081]$. Panel A presents estimates based on a 0.05 bandwidth, and Panel B presents estimates based on a 0.025 bandwidth. The table includes results from both linear and quadratic models, with either a rectangular or a triangular kernel weight function. Controls include individuals' gender, race, age, and an indicator of whether the BAC test was conducted in a traffic accident. Counterfactual predictions of mean recidivism are calculated at the 0.079 BAC threshold and mean age of the respective populations, averaging over individuals' gender and race, as well as over whether the BAC testing was conducted in an accident. Heteroscedasticity-robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

based on models' extrapolation over the region of the donut (**Dowd, 2021**) and may be especially sensitive to the assumptions about the underlying functional form (**Consider justifying this using a simulation**). Using polynomials with different orders can ensure the robustness of the results over different functional assumptions.

The results from donut regression discontinuity models are presented in Table 3. According to the local linear regression model with a rectangular kernel function, receiving punishments at the threshold decreases recidivism by 2.6 percentage points and is statistically significant at the level of 0.01. According to the local second-order polynomial with a rectangular kernel function, receiving punishments at the threshold decreases recidivism by 1.4 percentage points and is statistically significant at the level of 0.1. These estimates are consistent across both bandwidths and across models with different types of kernel functions, except for not being statistically significant in quadratic models estimated with narrower bandwidth of 0.025. Therefore, the results given by donut regression discontinuity models are essentially identical with those given by ordinary models presents in the last section.

3 Discussion, Critique and Extension

Using a regression discontinuity design, the present study aims to replicate the findings of **Hansen (2015)** about the effect of punishments and sanctions on reducing repeat drunk driving. Specifically, the present study focuses on testing whether receiving punishments at the 0.08 Blood Alcohol Content (BAC) threshold reduces future recidivism of drunk driving with a local linear regression model. The findings show that receiving punishments at the threshold reduces recidivism by up to 2.4 percentage points ². Robustness test based on local second-order polynomials show that receiving punishments at the threshold reduces recidivism by up to 1.6 percentage points ³. These estimates are consistent with those given by models that take into account non-random heaping in the data. The findings of the present study successfully replicate the main results of **Hansen (2015)**, which gives an estimated decrease in recidivism of up to 2 percentage points due to receiving punishments at the 0.08 BAC threshold. The success of replication demonstrates the validity of the methodology and results of the original study.

Hansen (2015) also presents related results that the present study, due to its scope and the limitation of the data available, does not replicate. As a part of the strengths of the original study, these findings either establish the robustness of the main results under different assumptions or reveal the heterogeneity of the effect of receiving punishments.

In terms of the robustness of the main results, the original study estimates regression discontinuity models using polynomials up to the third order, three types of kernel functions, and based on a series of bandwidths that extend beyond 0.05 and 0.025. The results show that the point estimates are relatively consistent, except for those given by the local second-order polynomial and those estimated based on very small bandwidths.

The heterogeneity of the effects of punishments include the effects among different populations and the effect on different types of recidivism. First of all, the original study separates analyses for drivers with and without prior drunk driving experience, in addition to running regression models on the entire population of drivers. This allows the original study to find that the effect of receiving punishments at the threshold is more pronounced among drivers with previous drunk driving experience (by about 4.6 percentage points, which is 25.2 percent ⁴). Moreover, the study conducts sensitivity analyses using different measures of prior drunk driving experiences, including whether a driver has taken BAC tests, whether a driver has a previous BAC over 0.08 (and is considered a case of DUI), as well as whether a driver has a prior conviction. Across all measures, regression models give point estimates and confidence intervals that are largely similar to each other. This demonstrates the robustness of the differential effects of receiving punishments at the threshold on different populations of drivers. Second, the original study conducts a detailed analyses of the effect of receiving punishments on three types of recidivism, refusal of taking a BAC test, and the likelihood of being stopped by a police. It finds that receiving punishments decreases the likelihood of being stopped as well as recidivism of all kinds. The effect is the largest on recidivism in which a driver has a BAC values in

²The average estimated effect is 2.1 percentage points. The average estimate is calculated over models based on different bandwidths and with different kernel functions, assuming equal weights of estimates.

³The average estimated effect is 1.5 percentage points.

⁴The estimate of percentgae points is calculatied by taking the mean over the two estimates reported in Table 3 of Hansen (2015).

[0.08, 0.15].

Despite multiple strengths of the original study, it can still be improved in several aspects. First, although sensitivity analyses using different bandwidths in the original study establish the robustness of the results, bandwidths of 0.05 and 0.025 still seem to be arbitrary choices of the researcher. Bandwidth choice can be improved by using data-driven methods with a chosen principle. For example, Mean Squared Error (MSE) optimal bandwidth choice method is available and has been recently improved with a bias correction of point estimate for better inference (**Imbens and Kalyanaraman, 2012; Cattaneo and Vazquez-Bare, 2016**). Alternatively, when the goal is to construct optimal confidence intervals for inference, one can choose to use a bandwidth selection method that minimises coverage error (CE) probability (**Calonico et al., 2022**). Both MSE-optimal and CE-optimal bandwidths are based on the data and constructed according to an explicit desideratum, which makes the bandwidth choice more objective and transparent. Future research can and should make full use of these methods as they are readily available in packages of major statistical software (**Calonico et al., 2015; Calonico et al., 2017**).

Second, in light of the present study, local second-order polynomials consistently give lower estimates than linear models. In fact, this is also the finding from the robustness test of the original study (Appendix Table 5 of Hansen, 2015). This suggests that the estimates of the effect of receiving punishments at the threshold are somewhat dependent on the assumptions of the effect's functional form. Although it is difficult to pin down one reasonable functional assumption due to the complex nature of the effect of punishments, future research should give equal weights in presenting the results given by different models. This is especially important for such an estimand as the effect of punishments on repeat drunk driving that is closely associated with policy making, since even a small difference in the estimate can lead to big differences when scaled up.

Third, the use of linear models is undesirable when recidivism, the outcome variable, is a binary variable. In simple words, this is because a mathematical inconsistency exists between the estimand (which can only take the value of 0 and 1) and the potential outcome of a linear model (which can take the value of any real number). In practice, future research can utilise logistic regressions in combination with regression discontinuity design to estimate the effect of punishments on recidivism. This should not be a difficult task because recent research in econometrics has developed and programmed such methods based on the multinomial logit model, together with optimal bandwidth selection and bias correction techniques (**Xu, 2017**).

Fourth, the robustness checks using donut hole regression discontinuity models are limited considering the possible non-random heaping in the data. The original study used donut hole models primarily to deal with potential confounders, but non-random heaping might also exist in light of the present study. Although the original study runs the model with observations dropped within windows of different sizes and found relatively consistent results, this method could still give biased estimates if non-random heaping outside the windows was biasing the estimation (**Barreca et al., 2016**). Future studies can adopt a potential solution suggested by previous research, that is, to separate the analyses for non-heaped data and heap points (if any). Unbiased estimates can be derived by taking a population-weighted average of the two estimates based on the restricted sample sizes (**Barreca et al., 2016**).

Finally, the external and internal validity of the study can be improved in future extensions. To improve the external validity of the estimation, one can use data from regions other than the state of Washington. This is to establish that punishments are universally effective in reducing repeat drunk driving across regions and the effect is not caused by some other factors specific to the state of Washington. The internal validity of the causal relation can be improved by conducting a longitudinal field experiment. One can choose a region where laws criminalizing DUI are not yet enforced, and carry out an experiment by manipulating the magnitude of the punishments drivers will receive if found driving under the influence. Drivers with a BAC level just below a chosen threshold are randomly assigned to receive lighter punishments than those with a BAC level just above the threshold. This method can theoretically provide stronger evidence for the causal relation between receiving punishments and recidivism. However, there are also challenges to such designs. For example, methodologically speaking, it is hard to prevent spillover effect. Drivers may find out punishments vary among individuals, and the consequence of such an effect is essentially unpredictable. Ethically speaking, it also requires debate on how such an experiment adhere to the principle of transparency. Due to the scope of this paper, I will not expand further on these topics and will leave them to be addressed by future research.

4 Conclusion

The present study successfully replicates the main results of the study *Punishment and Deterrence: Evidence from Drunk Driving* (Hansen, 2015) about the effect of punishments and sanctions on reducing repeat drunk driving using a regression discontinuity design. Specifically, I find that receiving punishments at the 0.08 BAC threshold reduces recidivism by up to 2.4 percentage points according to local linear regression models, and the decrease is up to 1.6 percentage points according to local second-order polynomials. These results provide evidence for the validity of the original study. The original study also goes beyond these estimates and reveals heterogeneity of the effects. Despite these strengths, the study can still be improved in aspects such as bandwidth selection, the selection of appropriate models and robustness checks, the presentation of the results, as well as internal and external validity.