HEC MONTREAL

SEVERITY OF PUNITIVE SENTENCE AND CRIME DETERRENCE

by

Léo Hassaoui

Decio Coviello
HEC Montréal
Research Director

Thesis presented in view of obtaining
the degree of Master of Science (M.Sc.) in Applied Economics

April 2023

Les membres du jury qui ont évalué ce mémoire ont demandé des corrections mineures.

Table of content

Acknowledgment

I would like to express my sincere gratitude to HEC Montréal, particularly the Department of Economics, for providing me with the opportunity to pursue my master's degree at their esteemed institution. The faculty have contributed greatly to my academic and personal growth and their dedication to excellence and commitment to educating future leaders in the field of economics has been a true inspiration.

Furthermore, I would like to express my gratitude to the teachers who taught me during my bachelor's degree program, whether it was in Canada or abroad. Their knowledge and expertise laid the foundation for my academic pursuits and provided me with the necessary skills and tools to succeed in my master's program. Thank you all for your unwavering support and encouragement throughout my academic journey. Your contributions have been immeasurable, and I will forever be grateful for your impact on my life.

Finally, I would also like to thank all the professors who have taught me during my master's degree program. Their rigorous academic instruction and mentorship have greatly enriched my academic journey. I am especially grateful to Decio Coviello, the research director of my thesis, for his invaluable guidance and support throughout my graduate studies.

Résumé

L'objectif de ce mémoire de maîtrise est d'évaluer l'impact de la sévérité des peines punitives sur la dissuasion de la récidive chez les criminels canadiens à l'aide d'un modèle de discontinuité de la régression. L'étude examine les systèmes de justice pénale pour les mineurs et les adultes au Canada et étudie les concepts de dissuasion et de choix rationnel dans le contexte de la criminalité. La recherche s'appuie sur le modèle de criminalité de Becker, qui explique comment les individus prennent des décisions rationnelles dans le cadre d'une analyse coûts-avantages de la criminalité. Avec un seuil à 18 ans (c'est-à-dire que nous comparons une cohorte de jeunes de 17 ans à une cohorte d'adultes de 18 ans), les résultats indiquent un effet dissuasif statistiquement significatif de la sévérité sur la récidive. Selon les modèles utilisés, le fait d'être traité comme un adulte réduit la probabilité de récidive d'environ 17 à 23%. Dans l'ensemble, on peut conclure que les adultes ont une probabilité de récidive significativement plus faible que les mineurs, et qu'un traitement plus sévère a un impact significatif sur la réduction de la probabilité de récidive.

Abstract

The objective of this master's thesis is to assess the impact of punitive sentence severity on recidivism deterrence among Canadian criminals using a regression discontinuity design. The study examines the criminal justice systems for juveniles and adults in Canada and investigates the concepts of deterrence and rational choice in the context of crime. The research draws on Becker's model of crime, which explains how individuals make rational decisions in a cost-benefit analysis of crime. With a cut-off at 18 years old (i.e., we compare a cohort of 17 years old youths with a cohort of 18 years old adults), the results indicate a statistically significant deterrence effect of severity on recidivism. According to the models used, being treated as an adult reduces the probability of reoffending by about 17 to 23%. Overall, it can be concluded that adults have a significantly lower likelihood of recidivism than juveniles, and that harsher treatment has a significant impact on reducing the probability of reoffending.

List of figures

List of tables

List of abbreviations & definitions

ACCS Adult Criminal Court Survey

CAD Canadian dollars

CCJS Canadian Centre for Justice Statistics

CJARS Criminal Justice Administrative Record System

FDLE Florida Department of Law Enforcement

ICCS Integrated Criminal Court Survey

MLE Maximum likelihood estimator

OLS Ordinary least square

PSC Public Safety Canada

RDD Regression discontinuity design

YCS Youth Court Survey

1. Introduction

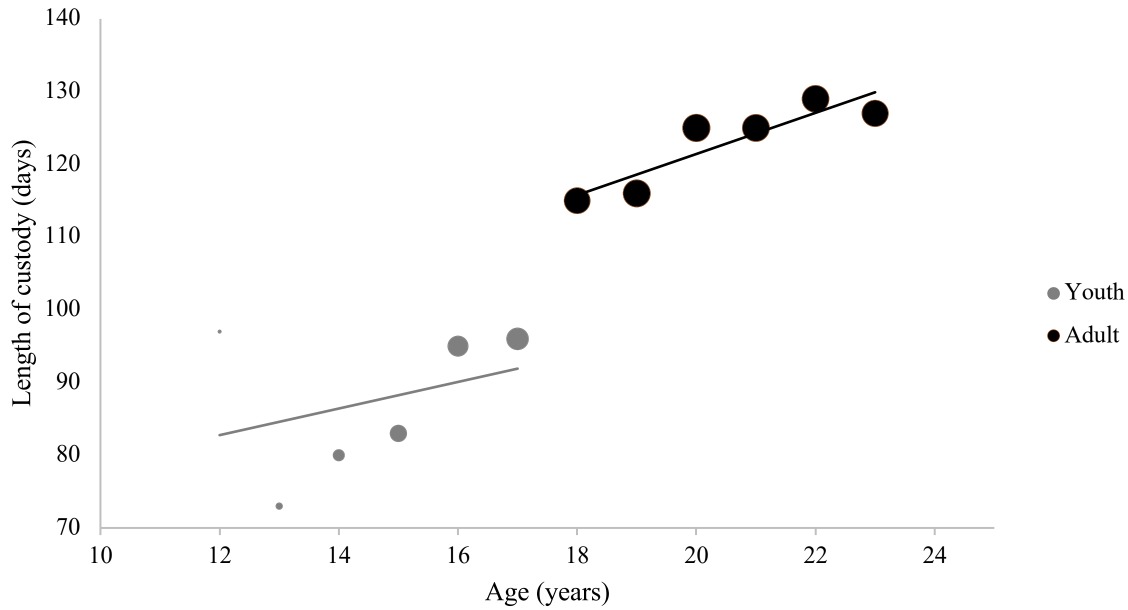
In this thesis we aim to investigate the impact of punitive sentence severity on recidivism deterrence in Canada using a regression discontinuity design. The criminal justice systems of many Western nations, including Canada, distinguish between juveniles and adults when prosecuting and sentencing offenders, primarily based on age. Violent crimes, such as homicide, assault, and robbery, have been identified by Public Safety Canada (PSC) as the costliest offenses when breaking down the cost of each crime. The Canadian Social Survey (CSS) estimates that the total cost of crime in Canada in 2009 was approximately 93.2 billion dollars, including all related expenses. Consequently, policymakers face a crucial decision-making process when it comes to crime control policies, as they have the potential to bring significant benefits to society. One of the most frequently cited methods for reducing crime is increasing the severity of punishment, as it is thought to act as a deterrent to criminal behavior.

In economics, the theory of rational choice posits that individuals make decisions based on a cost-benefit analysis, weighing the potential benefits of their actions against the potential costs. In the context of crime, this means that individuals may weigh the potential benefits of committing a crime against the potential costs, such as the risk of getting caught and punished, where the latter might deter individuals from committing a crime. In fact, deterrence refers to the idea that the threat of punishment can prevent individuals from engaging in criminal behavior by increasing the perceived costs of the action. When the punishment for a crime becomes more severe, the perceived cost of engaging in criminal behavior increases, making it less attractive for individuals to commit the said crime. This means that a more severe punishment, such as a sentence, can serve as a deterrent to crime by altering the cost-benefit analysis of potential offenders and reducing their incentive to engage in criminal behavior. Becker's model of crime, which was first introduced in 1968, provides a detailed explanation of how individuals make rational decisions in a cost-benefit analysis of crime. This will be further explored in Section 2.1.

In Canada, it is a well-established fact that adults are subjected to harsher treatment than youth, with the age of 18 serving as the dividing line between juvenile and adult status.

Consequently, a crime committed just before an individual's 18^{th} birthday is likely to receive different treatment, and possibly a less severe sentence, than one committed just after. This discrepancy in treatment is depicted in Figure I, which shows the relationship between age and the length of custody.

Figure I. Length of Custody[1] in Function of Age



The difference in the average length of custody at the age of 18 is significant. The average length of custody for youth aged 12 to 17 is 87 days, while for adults aged 19 to 23, it is 123 days. The figure displays the average length of custody per year over a six-year span, and the size of the dots represents the number of observations in each group. For example, the 12-year-old group, with a sample size of 825 observations, has an average of 97 days in custody. On the other hand, the 18-year-old group has an average of 115 days in custody, with a sample size of 74,222 observations. Although there are fewer observations for the youth group than the adult group, the clear discontinuity around the age of 18 implies that adults

[1] Length of custody is defined as the time which remains on a custodial sentence, not the total length of the sentence ordered by the judge because custodial sentence lengths reported to the survey exclude time spent in custody prior to sentencing and/or the amount of credit awarded for time spent in pre-trial detention (remand).

receive harsher sentences on average, assuming that longer custody time is a more severe punishment.

However, it is still unclear whether this increased severity is enough to deter criminals from reoffending since previous literature investigating the deterrent effect of punishment has yielded inconclusive results. In fact, past research has attempted to quantify and isolate a causal effect of punitive sentencing using different identification strategies and empirical models, but the results have been mixed.

Although the articles mentioned in Section 1.1 include the concept of time and duration, none of them explicitly use a duration model. The closest example is McCrary and Lee (2005), which uses a logit model to estimate the probability of committing a crime over a certain period, but not in a regression discontinuity design setting nor using a duration model. Our study is partly inspired by Lalive's (2006) study of the duration of unemployment benefits using sharp discontinuities in treatment assignment at age 50, and will use a similar approach to study the recidivism hazard at the treatment cut-off of the 18th birthday. This different strategy could provide a better understanding of the causal effect of the severity of punitive sentencing on recidivism. Therefore, this project aims to make two main contributions.

Firstly, we will attempt to replicate the methods used by McCrary and Lee (2005), Mueller-Smith, Pyle, and Walker (2022), and Lovett and Xue (2018) by computing recidivism probabilities and volumes around the age cut-off. However, we will expand our study to include all crimes that result in incarceration, not only violent crimes. Fortunately, our dataset allows us to examine more individuals, a different time period, and a different geographical region.

Secondly, we will enhance our base model by implementing a survival analysis which involves considering the duration of abstinence from crime (time elapsed before recidivism). The identification strategy will remain the same as in our base model, but our empirical approach will mimic the one used by Lalive (2006). We believe a duration model can provide more information and further validate whether severity of punitive sentencing can deter crime, as recidivism is a time-to-event data, and a duration model accounts for the timing of events and censored data.

The main results suggest that treating offenders more severely, i.e., as adults, has a significant impact on reducing the likelihood of recidivism. The findings support Becker's theory of crime, which suggests that the severity of punishment plays a key role in deterring criminal behaviour. However, the study's use of low-frequency data is a major issue, and future research could build on these findings by using higher-frequency data to further explore the observed impact of court sentencing on recidivism rates. Despite this limitation, we believe that this study provides valuable insights into effective strategies for preventing criminal behaviour.

1.1 Related Literature

McCrary and Lee (2005) used the administrative database maintained by the Florida Department of Law Enforcement (FDLE) to test the hypothesis that a more severe punishment can deter future criminal activities by increasing the perceived cost of crime. They based their hypothesis on Becker's (1968) economic model of crime, which views criminal behavior as the outcome of a cost-benefit analysis. To quantify the causal effect of punishment on crime, they employed a regression discontinuity design and used logit regression to examine the recidivism propensity of criminals one week before and one week after turning 18. The researchers exploited the fact that criminals are treated more harshly overnight, as they are subject to the adult criminal justice system rather than the youth criminal justice system. By comparing the two groups separated by only one week, they assumed that the treatment group and the control group were nearly perfect counterfactuals. They expected to find a discontinuous drop in the probability of offense at the age of 18, as the cost of crime increases significantly due to the transition from the juvenile to the adult justice system. However, their analysis revealed a non-statistically significant effect of deterrence, indicating that criminals did not tend to recommit crimes one week after turning 18 any less than their counterparts one week prior. This finding suggests that the severity of punishment alone may not be sufficient to deter criminals from reoffending, despite the increase in perceived costs associated with the transition to the adult criminal justice system.

This paper is a significant and impactful contribution to the field as the authors are the first to use a quasi-experiment generated by criminal law to compute recidivism rates in function of severity. That said, to better understand the impact of severity on recidivism, this paper will use McCrary and Lee’s strategy using a countrywide data set containing criminal data from all around Canada, instead of only one US state. Furthermore, we use a larger data set on arrests over a more recent and longer period. However, since we do not have the exact date of birth, the validity of our results can be easily questioned as the underlying assumptions of the regression discontinuity design that we use might not be respected, which we will discuss further in Section 5.

Alternatively, Mueller-Smith, Pyle and Walker (2022) find that adult prosecution reduces future criminal charges over 5 years by 20%. These authors use a different database that is from the Criminal Justice Administrative Record System (CJARS), the extent of their studies focuses on the state of Michigan. Again, using the discrete age of majority rule as their identification strategy, they use a regression discontinuity design to isolate a causal effect of severity of punitive sentence on crime deterrence. More specifically, they develop a novel econometric framework that combines standard regression discontinuity methods with predictive machine learning models to identify mechanism-specific treatment effects. In comparison to Mueller-Smith, Pyle, and Walker's study, our research has a notable advantage concerning the duration of our dataset. Specifically, we can examine recidivism over a period of up to 20 years, instead of just 5 years.

Similarly, Lovett and Xue (2018) have found that increasing the severity of punishment for criminals after they turn 18 can deter violent crimes by 10-12%. They also discovered that certain demographic subgroups are more responsive to sanctions than others. For instance, female, white, and Asian offenders have a lower recidivism rate when charged as adults. The authors draw their idea from Becker's theory of crime (1968), which suggests that agents choose an optimal level of crime by weighing the expected benefits against the expected costs of committing a crime. To obtain their results, they use administrative data from California and implement a regression discontinuity design by using the age cut-off point at the day level. Once again, our study has the advantage of using country wide data instead of a single state.

Finally, using data regarding drug arrests from the Research and Evaluation Division of the Chicago Police Department, Loeffler and Grunwald (2015) use a slightly different identification strategy by examining a sample that includes only non-transfer-eligible juveniles at the 17th birthday cut-off point. They employ a general linear model to estimate the log odds of rearrest and use a regression discontinuity design to assess the likelihood of criminal recidivism. They find that processing juveniles as adults slightly reduces the probability of recidivism by 3-5%. However, non-transfer-eligible juveniles account for a small proportion of arrestees, whereas our data set contains considerably more observations around the cut-off of 18 years old.

2. Conceptual Framework

2.1 Economic Theory of Crime

Firstly, it is important to understand how this research project fits into economic theory and how it differs from criminological theory. The economic and criminological approaches to criminal behavior differ in several ways, which will be discussed in this paper. Criminological literature can be broadly categorized into three main branches. The first field of research examines biological causes of crime, such as brain abnormalities or hormone imbalances. The second subset of criminological literature suggests that individuals turn to crime when they cannot achieve their goals through legal means. The third subfield of criminological literature focuses on social interactions to understand if criminal behavior is learned or socially transmitted. In contrast, the economic approach assumes that these three subfields account for the baseline level of crime in a cost-benefit analysis of criminal behavior. Rather than discounting the importance of these elements, the economic approach acknowledges their role in shaping criminal behavior. The economic approach assumes that people will weigh the costs and benefits of criminal activity against those of lawful pursuits and will only commit a crime if the benefits outweigh the costs. Therefore, the economic model of crime covers most criminological justifications for criminal behavior.

Next, let's discuss how the economic theory of crime relates to recidivism. Recidivism can sometimes be attributed to unpredictable behavior, lack of self-control, or evidence that the deterrence model is ineffective. For instance, criminals who re-offend may perceive that the benefits of doing so outweigh the costs. Additionally, if the initial cost of committing a crime was reasonable (i.e., the sentence was reasonable), many offenders will remain motivated to commit crimes after serving their time in jail. However, the likelihood or severity of punishment may increase in the estimations of certain offenders who have already been convicted or imprisoned. These criminals may be less inclined to commit another crime because the costs of doing so are higher and/or more accurately estimated than the costs associated with their first offense. Thus, the decision to reoffend depends on past punishment and the perception of potential future punishments, leading us to our third point.

Thirdly, the economic theory of crime highlights how individuals' discounting of future rewards and punishments can influence their decisions regarding illegal actions. Since crime often brings immediate benefits, while punishment may be delayed, individuals with a high discount rate are more likely to engage in illegal behavior, giving greater weight to present benefits over future costs. Economic theory suggests that rational individuals adapt their behavior based on changes in the expected costs and benefits of illegal conduct. Becker's work, in particular, emphasizes how choices can be understood within an economic framework, even if they are influenced by subjective beliefs. Thus, if the expected costs of committing or recommitting a crime increase significantly or if the expected benefits decrease, individuals are likely to be deterred from engaging in criminal behavior, even if their estimation of costs and benefits is imperfect.

Using Becker's base economic model of crime, the expected utility function of a criminal is modelled as follows:

$$E[U_i(Y,a)] = pU_i(Y+a) + (1-p)U_i(Y) \quad (2.1)$$ [2]

Where $U_i(Y,a)$ is the von Neumann-Morgenstern utility function of criminal i, Y is the payoff of a given crime ($Y > 0$), a is the fine of a given crime ($a < 0, a \leq Y$) and p is the probability of punishment ($0 < p < 1$).

[2] The following model can be derived if we expend the base model of Becker's expected utility of criminals by considering the discount factor β over time t ($0 < \beta < 1$):

$$\sum_{t=0}^{T} \beta_i^t \, E[U_i(Y_t, a_t)] = \sum_{t=0}^{T} \beta_i^t \left[pU_{ij}(Y_t + a_t) + (1-p)U_i(Y_t) \right]$$

Let's also note that Becker empirically observed that:

$$\eta_p^{U_i(Y,a)} \equiv \frac{\partial E[U_i(Y,a)]}{\partial p}\frac{p}{U_i(Y,a)} = [U_i(Y) + U_i(Y+a)]\frac{p}{E[U_i(Y,a)]} \tag{2.2}$$

$$>$$

$$\eta_a^{U_i(Y,a)} \equiv \frac{\partial E[U_i(Y,a)]}{\partial a}\frac{a}{U_i(Y,a)} = pU'_i(Y+a)\frac{f}{E[U_i(Y,a)]} \tag{2.3}$$

Meaning that the elasticity of expected utility with respect to the probability of punishment ($\eta_p^{U_i(Y,a)}$) is greater than the elasticity of expected utility with respect to the fine ($\eta_a^{U_i(Y,a)}$), implying that criminals respond more to an increase in probability of punishment than to an increase in severity of punishment. However, it is very important to note than even though $\eta_p^{U_i(Y,a)} > \eta_a^{U_i(Y,a)}$, the elasticity of expected utility with respect to the fine is negative ($\eta_a^{U_i(Y,a)} < 0$), implying that an increase in severity of punishment decreases the expected utility of criminals. We can therefore logically deduct that the elasticity of expected utility with respect to the probability is also negative, although a little more negative ($\eta_a^{U_i(Y,a)} < \eta_p^{U_i(Y,a)} < 0$).

These observations are particularly important to this study for three main reasons:

1. Expecting a high probability of punishment may disincentivize criminals to (re)commit a crime, such that $(\mathbf{1} - \boldsymbol{p})U_i(Y) - \boldsymbol{p}U_i(Y+a) < 0$. In the case of a first crime, this could be explained by being a risk averse criminal. In the case of a second crime, it could be explained by some sort of realization that getting caught is a strong possibility.

2. Expecting a high fine may disincentivize criminals to (re)commit a crime, such that $(1-p)\boldsymbol{U_i(Y)} - p\boldsymbol{U_i(Y+a)} < 0$. In the case of a first crime, this could be explained by a fear of potential fine. In the case of a second crime, this could be explained by a realization of the costs of committing a crime following a first arrest and therefore a first fine.

3. If criminals respond more to an increase in probability of getting caught than to an increase in fine, criminals must be risk preferring. In fact, if the criminal is risk averse, we have that $U'' < 0$ and that therefore, the relation between equation (1) and (2) doesn't hold since if $\eta_p^{U_i(Y,a)} > \eta_a^{U_i(Y,a)}$, then $[U_i(Y) + U_i(Y+a)]\frac{p}{E[U_i(Y,a)]} > pU'_i(Y+a)\frac{f}{E[U_i(Y,a)]}$, or $\frac{[U_i(Y)+U_i(Y+a)]}{f} > U'_i(Y+a)$, which is true if and only if U_i is risk preferring ($U'' > 0$). In the case that criminals are risk averse, it should then be that they would react more to an increase in expected severity of punishment than in the increase in certainty of punishment.

The third point becomes particularly interesting, in the sense that it could explain why past literature's results have been so mitigated – higher severity could lead to disincentivize criminals to recommit crime, although it could not, depending on the risk aversion of criminals.

Economic theory therefore states that a high punishment (fine) given after a first crime should deter recidivism by increasing the expected future cost, if and only if the criminal is rational (maximizes his utility) and risk averse (correlate positively punishment severity and cost of crime).

2.2 Identification Strategy

Like Lalive (2006), Lovett and Xue (2018), McCrary and Lee (2005), and Mueller-Smith, Pyle, and Walker (2022), we will use the age threshold that separates the juvenile and adult justice systems. When a criminal turns 18, the entire criminal justice system and the way defendants are handled changes drastically. By comparing individuals who are nearly identical but are on different sides of the age cut-off, we can obtain credible estimates of the causal effects for a specific subpopulation. In other words, a criminal who is 18.01-year-old can serve as a near-perfect counterfactual for a 17.99-year-old criminal, assuming that they are identical apart from the fact that the latter is handled in the youth criminal justice system and the former in the adult criminal justice system.

Let's say that we have two potential outcomes $Y_i(t = 0)$ and $Y_i(t = 1)$, then the causal effect is as follows: $Y_i(1) - Y_i(0)$, where $t \in \{0,1\}$ is the treatment, with $t = 0$ being when no treatment is received and $t = 1$ being when a treatment is received. In our study, the treatment around the cut-off point is in fact whether an individual is treated as an adult ($t = 1$) or not ($t = 0$). Then, we have:

$$Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i) = \begin{cases} Y_i(1) \ if \ a_i \geq c \\ Y_i(0) \ if \ a_i < c \end{cases} \quad (2.4)$$

Where T_i is a deterministic function of one of the covariates, the treatment-determining variable age a, that is:

$$T_i = \mathbf{1}\{a_i \geq c\} \quad (2.5)$$

This is true in our case since we use a sharp regression discontinuity design and that therefore once we know a_i, we know T_i.

One of the fundamental problems of causal inference occurs because we can only observe the outcome under control ($Y_i(0)$) for units whose score is below the cut-off c and we can only observe the outcome under treatment ($Y_i(1)$) for units whose score is above the cut-off c. That is, there is no value of X_i at which we observe both treatment and control observations. From this, a sharp regression discontinuity design can be written as follows:

$$\lim_{a \to c^+} E[Y_i(t=1)|T_i = 1, a_i = a] - \lim_{a \to c^-} E[Y_i(t=0)|T_i = 0, a_i = a] \tag{2.6}$$

Or, since we know (5), we get:

$$\lim_{a \to c^+} E[Y_i(1)|a_i = a] - \lim_{a \to c^-} E[Y_i(0)|a_i = a] \tag{2.7}$$

And under the assumption that $E(Y_i(0) \mid X_i = x)$ and $E(Y_i(1) \mid a_i = a)$ are continuous in a, we can derive the average causal effect of the treatment at the discontinuity point:

$$\theta_{SRDD} = E[Y_i(1) - Y_i(0)|a_i = c] \tag{2.8}$$

Where θ_{SRDD} is the average causal effect derived from the sharp regression discontinuity design, t is the treatment (i.e., adult criminal system), c is the cut-off point (i.e., 18 years old) and a is the age.

However, it is very important to note that this average causal effect can be interpreted without bias if and only if the continuity of conditional regression functions stands, that is:

$$E[Y_i(0)|a_i = a] \; and \; E[Y_i(1)|a_i = a] \; are \; continuous \; in \; a \tag{2.9}$$

In fine, the robustness and reliability of our results depend on the following key assumptions;

A1. **The discontinuity is sharp at the cut off: the probability of assignment jumps from 0 to 1 at cut-off.** Since the criminal justice system treats individuals under the age of 18 as juvenile and above 18 as adults, this assumption is respected – the probability of being treated in the adult criminal justice system jumps from 0 to 1 at the age of majority. That is, the probability of treatment is discontinuous at the cut-off.

A2. **Assignment occurs through a known and measured deterministic decision rule.** Recalling equation 2.5 an assuming A1, once we know a_i, we know T_i. That is, once we know the age, we know if the individual will be treated in the adult criminal justice system or not.

A3. **There is local continuity: subjects just above and below the cut-off have similar potential outcome.** Recalling equation 2.9 and given that the data used in this study is of low frequency, (i.e., yearly observation grouping), A3 is a very strong assumption. If the data were of high frequency, we could assume that "all other factors" are similar when examining the treatment group (adults) against the control group (youths). However, due to the low frequency of the data used in this study, it is possible that many factors other than being treated as an adult could have affected criminal behavior. For example, psychological changes, finishing high school, or starting a new job could be confounding factors. Therefore, the assumption that there is local continuity, where subjects just above and below the cut-off have similar potential outcomes, might not be respected in our analysis. This limitation will be further discussed in Section 5.

2.3 Empirical Strategy

We will use two empirical models to implement our identification strategies and quantify the causal effect of punishment severity on recidivism. The first model aims to replicate the study by McCrary and Lee (2005) using our dataset, which, as a reminder, includes more observations, different control variables, and a longer time period. By utilizing more comprehensive panel data, this will allow us to further understand the deterrence effect of punitive sentencing. The second model will exploit the duration component of recidivism, which we believe can provide more information than with a logistic approach since recidivism is a time-to-event data.

2.3.1 Logistic Model

The approach used to compute our results is quite straightforward. To implement our identification strategy with regards to our logistic model, we first need to identify all youths who have committed a crime before the age of 16. For instance, if a criminal had committed a crime at the age of 14, we would keep all observations regarding that person and start tracking them from their 16th birthday onwards. Once this is done, we flag all criminals that have committed a first crime at a given year between the age of 16 to 19. That is, our whole sample contains criminals between 16 and 19 years old only, with some criminals having committed their first crime before 16 and a second crime between 16 and 19, as well as criminals having committed their first crime at the age of 16, 17, 18 or 19 respectively. Finally, we compute the probability of recommitting a crime for each subsequent year after turning 16, until we reach the age threshold, and just pass it. That is, we compute the probability of recidivism for the whole sample by looking at the proportion of individuals who have recommitted a crime at 16, 17, 18 and 19 compared to the proportion of individuals who haven't recommitted a crime at the same given age. Excluding youths who have committed a crime before the age of 16 but never after has the advantage of observing only individuals who are more likely to understand the implications and differences between the juvenile and adult justice systems. This approach is very similar to the one used by McCrary and Lee (2005).

2.3.1.1 Econometric approach

First, the following logistic regression will be estimated to compute the probability of recidivism around the cut-off age of 18 years old:

$$\log\left(\frac{p(a_i, \boldsymbol{X_i})}{1 - p(a_i, \boldsymbol{X_i})}\right) = \alpha + \theta\mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'_i\beta} \tag{2.10}$$

Or;

$$\left(\frac{p(a_i, \boldsymbol{X_i})}{1 - p(a_i, \boldsymbol{X_i})}\right) = e^{\alpha + \theta\mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'_i\beta}} \tag{2.11}$$

Or;

$$p(a_i, \boldsymbol{X_i}) = \frac{e^{\alpha + \theta\mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'_i\beta}}}{1 + e^{\alpha + \theta\mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'_i\beta}}} \tag{2.12}$$

Where p is the probability that criminal i has recommitted a crime ($1 - p$ being the inverse probability), θ is the average causal effect, estimated from a dummy variable taking the value 1 if criminal i is 18 years of age, or older ($a_i \geq 18$), and $\boldsymbol{X_i}$ is a vector of control variables for criminal i.

2.3.1.2 Estimation strategy

We now need to empirically estimate the parameters of equation 2.12. To do so, we will use the maximum likelihood estimator (MLE) approach. Let's first define the general likelihood function, which will also be useful when the second estimation strategy will be discussed:

$$\mathcal{L}(\pi) = \prod_{i=1}^{n} f(y_i|\pi) \tag{2.13}$$

Where the goal is to find values for the parameters π such that the likelihood function 2.13 is maximized, that is:

$$\hat{\pi} = \max_{\pi} \mathcal{L}(\pi)$$

More specifically, for the case of our logit model and by assuming a Bernoulli distribution, our general parameter π becomes a probability parameter p and our likelihood function 2.13 becomes:

$$\mathcal{L}(p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} \qquad (2.14)^3$$

But, because of equation 2.11 and the fact that $1 - p = 1 - p(a_i, \boldsymbol{X'}_{\boldsymbol{i}}) = \frac{1}{1+e^{\alpha+\theta\mathbf{1}\{a_i\geq 18\}+\boldsymbol{X'}_{\boldsymbol{i}}\boldsymbol{\beta}}}$, we can re-write equation 2.14 as follows:

$$\mathcal{L}(\alpha, \theta, \boldsymbol{\beta}) = \prod_{i=1}^{n}\left[\left(e^{\alpha+\theta\mathbf{1}\{a_i\geq 18\}+\boldsymbol{X'}_{i}\boldsymbol{\beta}}\right)^{y_i}\left(\frac{1}{1+e^{\alpha+\theta\mathbf{1}\{a_i\geq 18\}+\boldsymbol{X}'_{i}\boldsymbol{\beta}}}\right)\right] \qquad (2.15)$$

Where we end-up with the maximization problem, simplified by taking the log-likelihood, which is defined as follows:

$$\max_{\alpha,\theta,\boldsymbol{\beta}} \log \mathcal{L}(\alpha, \theta, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i\left(e^{\alpha+\theta\mathbf{1}\{a_i\geq 18\}+\boldsymbol{X}'_{i}\boldsymbol{\beta}}\right) - \sum_{i=1}^{n}\log\left(1+e^{\alpha+\theta\mathbf{1}\{a_i\geq 18\}+\boldsymbol{X}'_{i}\boldsymbol{\beta}}\right) \qquad (2.16)$$

Where we solve for the parameters $\alpha, \theta, \boldsymbol{\beta}$ such that it maximizes the likelihood function 2.16 on the observed values for $y_i = \{0,1\}$.

2.3.2 DURATION MODEL

As for the duration model, the approach differs from the one used for the logistic model. The first step is to identify criminals who committed their first crime at a given age. Then, we follow this cohort of criminals over time and track how long it took them to recommit a crime, if they did. We use individuals who committed their first crime because we have low frequency data, and it is a simple way of controlling for potential psychological effects on deterrence or recidivism. That is, we want to compare similar groups of individuals, meaning that if we were to compare recidivism rates between youths who have committed dozens of crimes and new criminals around the 18-year-old threshold, then the deterrence effect could

[3] From there on forward, we parametrize $p = p(a_i, \boldsymbol{X}_i)$. And, since we know that $p \in [0,1]$ and $a_i, \boldsymbol{X}_i \in \mathbb{R}^k$, our function $p(a_i, \boldsymbol{X}_i)$ is such that $p(a_i, \boldsymbol{X}_i) : \mathbb{R}^k \to [0,1]$. That way, we can use a vector of explanatory variables $(a_i, \boldsymbol{X}_i)$ of dimensions $1 * K$ associated to each criminal i.

be due to factors other than being treated as an adult. The objective here is to see whether being treated more severely can reduce the instantaneous rate of recidivism at any given time.

2.3.2.1 Econometric approach

For the second empirical model (i.e., duration model), the following regression will be estimated to compute the duration before recidivism around the cut-off age of 18 years old.

$$\log h(a_i, \boldsymbol{X_i}) = \alpha + \theta \mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'_i\beta}$$

$$h(a_i, \boldsymbol{X_i}) = e^{\alpha + \theta \mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'_i\beta}} \tag{2.17}$$

The use of an exponential distribution on our hazard function will be further discussed in the next section. All the parameters from equation 2.17 are defined identically as in 2.12, apart from $h(a_i, \boldsymbol{X_i})$ which is the hazard function of criminal i.

2.3.2.2 Estimation strategy

To estimate the parameters of our regression, we will use the MLE approach. It is important to note that because we have observations in our data set where some criminals have not recommitted a crime (no recidivism), we need to model a maximization likelihood problem considering right censoring. More specifically, let's suppose for a given criminal i, we have a survivor function $S(t) = 1 - G(t)$ that follows a distribution $g(t)$ with a hazard function $h(t)$, we then get for a criminal that has recommitted a crime at time t the following likelihood function:

$$\mathcal{L}(G) = g(t) = \big(1 - G(t)\big)h(t) = S(t)h(t) \tag{2.18}$$

On the other hand, if the criminal hasn't recommitted a crime after time t (right censored), we get the following likelihood function:

$$\mathcal{L}(G) = \big(1 - G(t)\big) = S(t) \tag{2.19}$$

Which, by combining equation 2.18 and equation 2.19 for all criminals and using the same principles as in the general likelihood function 2.13 (i.e., the general likelihood parameter $\pi \equiv G$), leads us to the following likelihood:

$$\mathcal{L}(G) = \prod_{i=0}^{n} g(t_i)^{c_i}\big(1 - G(t_i)\big)^{1-c_i} \tag{2.20}$$

Where $c_i = \begin{cases} 1 \text{ if criminal } i \text{ has recommitted a crime (recivism)} \\ 0 \text{ if criminal } i \text{ has not recommitted a crime (no recidivism)} \end{cases}$

Which is also often rewritten as:

$$\mathcal{L}(G) = \prod_{i=0}^{n} h(t_i)^{c_i} S(t_i) \qquad (2.21)^4$$

However, for the case of our duration model (i.e., by assuming an exponential distribution as per equation 2.17), the likelihood function incorporates our assumption through the following substitutions:

1. $g(t_i) = \lambda e^{-\lambda t_i}\ where\ \lambda > 0$
2. $h = \lambda$
3. $S(t_i) = \frac{g(t_i)}{h(t_i)} = e^{-\lambda t_i}$

[4] Where $h(t)$ is the instantaneous rate of occurrence of a given event (or hazard function), that is:

$$h(t) = \lim_{dt \to 0} \frac{\Pr\{t \le T < t + dt | T \ge t\}}{dt} = \lim_{dt \to 0} \frac{\frac{g(t)dt}{S(t)}}{dt} = \frac{g(t)}{S(t)}$$

The numerator being the conditional probability that the event will occur at T in the interval $(t, t + dt)$, given it hasn't already occurred. For a small dt, the numerator can also be written as the ratio of the joint probability that T is in the interval $(t, t + dt)$ **and** that $T \ge t$ (i.e., $g(t)dt$), to the probability of $T \ge t$ (i.e., $S(t)$, from the definition of the survival function; $S(t) = \Pr\{T \ge t\} = 1 - G(t)$). The denominator ($dt$) is the width of the interval. Therefore, from the relation $g(t) = S(t)h(t)$ and $1 - G(t) = S(t)$, we have that $g(t)^c\big(1 - G(t)\big)^{1-c}$ can be rewritten $\big(S(t)h(t)\big)^c(1 - S(t))^{1-c}$, and therefore can be simplified as $h(t)^c S(t)$.

It is important to note that, as discussed in Section 2.3.2.1, we use an exponential survival distribution because we cannot affirm that the longer you spend outside of jail the higher the probability will be to (or not to) recommit crime, which is what a Weibull distribution would imply, for instance. Instead, we assume that there is no correlation between time spent outside of jail post-release, and recidivism. That is, the probability of recidivism for criminals is constant over time (i.e., no time memory), which is why $h = \lambda$. Therefore, from equation 2.20, we end-up with the following maximization problem:

$$\max_{\lambda} \mathcal{L}(\lambda) = \prod_{i=0}^{n} (\lambda e^{-\lambda t_i})^{c_i} \left(e^{-\lambda t_i}\right)^{1-c_i} \tag{2.22}$$

Or with the following, when simplifying by taking the log-likelihood:

$$\max_{\lambda} \log \mathcal{L}(\lambda) = \left(\sum_{i=1}^{n} c_i\right) \log \lambda - \lambda \sum_{i=1}^{n} t_i \tag{2.23}$$

However, to be able to solve for the parameters $\alpha, \theta, \boldsymbol{\beta}$ such that it maximizes the likelihood function 2.20 on the observed values, we need to parametrize our model, to do so we can directly model one of the parameters of the distribution, (i.e., λ in the case of the exponential distribution) such that:

$$\lambda = e^{\alpha + \theta \mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'}_i \boldsymbol{\beta}} \tag{2.24}$$

We can therefore expand to a log likelihood maximization problem specific to our duration analysis with an exponential distribution, that is:

$$\max_{\alpha,\theta,\boldsymbol{\beta}} \log \mathcal{L}(\alpha, \theta, \boldsymbol{\beta}) = \left(\sum_{i=1}^{n} c_i\right) e^{\alpha + \theta \mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'}_i \boldsymbol{\beta}} - e^{\alpha + \theta \mathbf{1}\{a_i \geq 18\} + \boldsymbol{X'}_i \boldsymbol{\beta}} \sum_{i=1}^{n} t_i \tag{2.25}$$

Where we can finally solve for the parameters $\alpha, \theta, \boldsymbol{\beta}$ such that it maximizes the log likelihood function on the observed values.

3. Descriptive Statistics

3.1 Data

We utilized data from the Integrated Criminal Court Survey (ICCS), which provides information on court cases involving Criminal Code and other federal statute offenses in Canadian courts, including characteristics of the cases and the accused persons. The data covers all recorded arrests in Canada from 1994 to 2014 and includes over 12 million observations. The micro-data for the ICCS is maintained by the Canadian Centre for Justice Statistics (CCJS), Statistics Canada, and the data collection strategy is designed to integrate the collection of adult and youth court data. The ICCS replaces both the legacy Adult Criminal Court Survey (ACCS) and the legacy Youth Court Survey (YCS). The data allows for breakdowns by crime category, gender, location, and more. Further details about our sample will be discussed in the next Section.

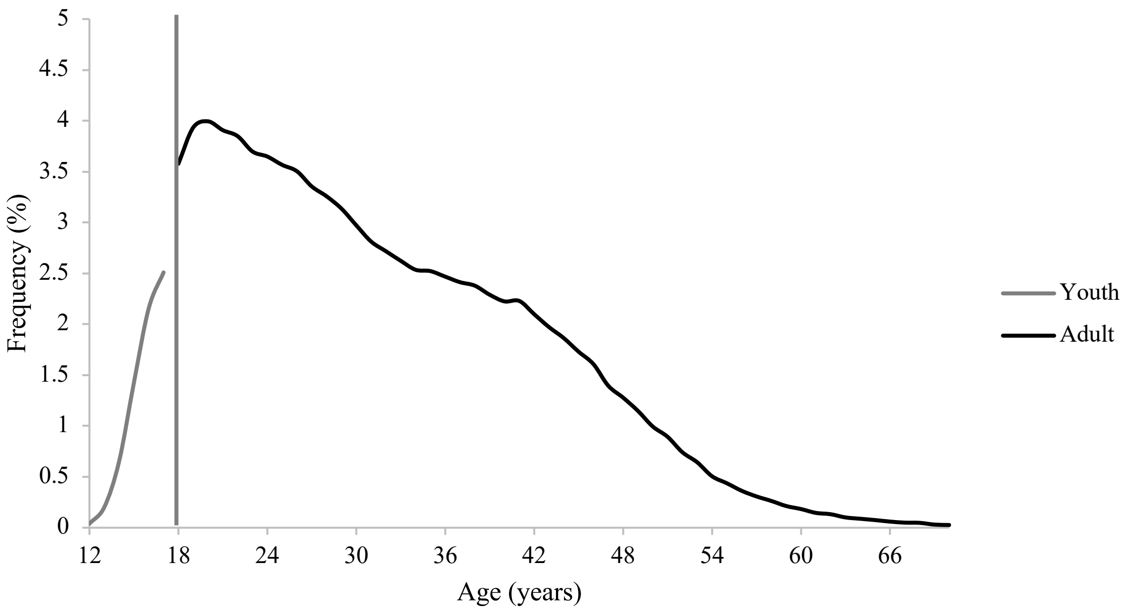
The data is highly suitable for our purposes because it covers both adults and juveniles and provides longitudinal data over a longer period than previous studies. We take advantage of this feature of the ICCS database to conduct a quasi-experiment, comparing the recidivism of suspects arrested and convicted just before and after their 18th birthday. With regards to our logistic model, we generate a recidivism variable by flagging criminals that have committed a crime before the age of 16 and that have recommitted a second crime at the age of 16, 17, 18 or 19 respectively. As for our duration model, we generate a recidivism variable by flagging criminals that were arrested at the age of 16, 17, 18 or 19 respectively and that have recommitted a crime at some point in time, where we finally generate a duration variable by computing the number of days between the first and second crime.

However, one issue we encountered with the data is the frequency of dates of birth, which limits us to pinpointing age at the time of the offense only in years, rather than days, weeks, or even months. We will address this issue in Section 5.

3.2 Samples

Essentially, we subdivided the original database into two sub-samples, and due to the large size of the database (more than 12 million observations), extensive data cleaning was necessary to prepare the data for analysis. The age distribution of criminals at the time of the offense in each of the two sub-samples is presented below.

Figure II. Age Distribution at the Time of Offence



Approximately 7% of all recorded offenses in our dataset were committed by individuals aged 12 to 17, which amounts to around 840,000 out of the total 12 million identified offenses. The distribution of offenses is heavily skewed towards younger individuals, with about 50% of all offenses being committed by those aged 30 or younger. It is worth noting that there is a significant jump in the frequency of crimes committed at the age of 18. This increase could potentially be explained by differences in police enforcement towards those under 18 or changes in socio-economic or psychological factors at the transition from 17 to 18 years old. While investigating this observation further would be interesting, the current project's focus is solely on identifying if a certain amount of crime is committed after reaching majority. However, investigating the underlying causes and potential solutions is a critical next step in addressing and reducing criminal activity.

3.2.1 FIRST SAMPLE – LOGISTIC MODEL

As discussed in Section 2.3.1, the first sample aggregates all individuals who committed a crime before age 16, and then tracks the sample over time starting from their 16th birthday. That is, all criminals who have committed at least one crime before the age of 16 are flagged, and only those flagged individuals between the ages of 16 and 19 are kept for the recidivism sample, while all other first-time offenders are included in the youth sample.

Table I. Summary Statistics – Offence, Logit

Variable	Youngster sample		Recidivism sample	
Male	0.88		0.85	
Age	17.81		16.40	
Case duration	248		253	
Charge duration	198		169	
Length of custody	105		101	
Number of individuals	**43,481**		**5,199**	
	Frequency	Length of custody	Frequency	Length of custody
Crime category				
Crime against the person	23.01%	193	27.20%	160
Crime against property	19.16%	104	21.33%	95
Administration of justice	30.31%	30	20.58%	62
Criminal Code traffic	3.48%	118	1.44%	115
Other	24.05%	114	29.45%	79
Sample size	**48,680**			

Note: Column one (Youngster sample) contains all observations from individuals that have committed a first crime between the age of 16, 17, 18 or 19 respectively. Column two (Recidivism sample) contains only observations from individuals between the age of 16 to 19 that have committed at least one crime before the age of 16 **and** that have recommitted for the first time a second crime between 16 to 19 years old.

Table I summarizes the data used for the first sample to calculate the probability of recidivism around the age of 18. Out of the 48,680 individuals who committed crimes between the ages of 16 and 19, the average age at the time of offense was around 18. On the other hand, the average age of the 5,199 individuals in the recidivism sample was approximately 16, indicating that those with a history of repeat offending tend to commit crimes at a younger age than those in the broader sample, which includes both first-time and repeat offenders. About 90% of offenses in both the younger and recidivism samples were committed by males, which is typical of criminal justice datasets. The recidivism sample had a shorter average charge duration, likely because these individuals already have a criminal record, leading to faster administrative processing and court proceedings.

The lower section of Table I provides a breakdown of the number of individuals and duration of custody by crime category. Crimes against persons and property (the most serious types of crime) increase significantly between the younger and recidivism samples. For example, crimes against persons accounted for about 23% of all offenses in the younger sample and slightly over 27% in the recidivism sample. Regarding these more serious crimes, it is interesting to note that individuals who reoffended tended to have a shorter length of custody. This could be due to the definition of the length of custody variable, which only includes the remaining time on a custodial sentence and does not consider time spent in pre-sentencing custody or credit for pre-trial detention. The shorter length of custody in the recidivism sample may also result from shorter charge and case durations. Finally, while administration of justice crimes were the most common crimes in the younger sample (excluding the "Other" category), in the recidivism sample, crimes against persons became the most prevalent crime category. This observation suggests that re-offenders tend to commit more violent crimes.

Table II. Descriptive Statistical Evidence – Recidivism Frequency Estimation

	16 years old		17 years old		18 years old		19 years old	
Recidivism	Yes	No	Yes	No	Yes	No	Yes	No
Number of individuals	3,866	7,296	756	7,583	385	14,557	192	14,045
Proportion of recidivism	52.99%		9.97%		2.64%		1.37%	

Note: All numbers under the “No” recidivism column represents the number of individuals that have committed their first crime at a given age, and never again.

Table II depicts the proportion of individuals with recidivism by age group. For example, in the age group of 17-year-olds, there were a total of 8,339 offenders, out of which 7,583 committed a crime for the first time in their lives at the age of 17. On the other hand, 756 individuals had reoffended for the first time at 17 after committing a crime between the ages of 12 to 15. It is not surprising to see a high proportion of recidivism at 16 years old, as we only count recidivism once. This means that if an individual committed a crime before the age of 16 and reoffended at 16, we stop following that individual at 17, 18, or 19 years old. Thus, the 16-year-old sample likely contains individuals with the highest delinquent tendencies, while the 17, 18, and 19-year-old samples comprise reasonably delinquent youths. With more frequent data, we could have validated this hypothesis. However, the low frequency data presents another challenge. While there is a significant drop in recidivism between the ages of 17 and 18 compared to the drop between ages 18 and 19 (-7.33% versus -1.27%), it is unclear whether this decrease is discontinuous or linear.

3.2.2 Second sample – Duration model

In Section 2.3.2, it was discussed that the second sample is comprised of individuals who committed their first crime at around the age of 18, and are tracked over time. This approach, which involves duration analysis, is more informative than the logistic approach as it considers the timing of events, accounts for censored data, and allows for more flexible modeling of hazard rates (e.g., exponential distribution or Weibull distribution). Since recidivism is a time-to-event data, our duration model can provide more information and further validate whether severity of punitive sentencing can deter crime.

Table III presents a summary of the second sample data used to calculate the duration before recidivism at around 18 years old. The analysis revealed that the average age at the time of the offense among the 8,305 individuals who committed crimes between ages 16 and 19 and never repeated the offense was approximately 18. In contrast, the average age among the 23,736 individuals in the recidivism sample was about 17, indicating that those with a history of repeat offending tend to have committed their first crime at a younger age than those with no recidivism. The proportion of males committing offenses was approximately 90% in both samples, which is typical of criminal justice datasets. The recidivism sample had a shorter average case and charge duration, likely due to faster administrative processing and court proceedings resulting from the criminal record of these individuals.

Table III. Summary Statistics – Offence, Duration

Variable	No recidivism sample		Recidivism sample	
Male	0.85		0.90	
Age	17.81		17.40	
Case duration	209		152	
Charge duration	189		113	
Length of custody	71		80	
Number of individuals	**8,305**		**23,736**	
	Frequency	Length of custody	Frequency	Length of custody
Crime category				
Crime against the person	30.10%	127	17.37%	127
Crime against property	13.88%	55	19.86%	93
Administration of justice	29.54%	14	35.38%	50
Criminal Code traffic	5.07%	89	2.63%	91
Other	21.41%	77	24.76%	79
Sample size	**32,041**			

Note: Column one (No recidivism sample) contains a cohort of individuals that committed a first crime at the age of 16, 17 or 19 respectively and never recommitted a crime after that. Column two (Recidivism sample) contains a cohort of individuals that committed a first crime at the age of 16,17 or 19 and recommitted a crime at some point later in life.

The lower section of Table III presents the number of individuals and length of custody by crime categories. The frequency of crimes against the person decreased significantly between the no recidivism and recidivism sample (-12.73%), while the frequency of crimes against property increased (+5.98%). These findings differ from those of the sample in Section 3.2.1, but this can be attributed to the differences in methodologies used to observe recidivism. Reoffenders tend to have a shorter length of custody, and administration of justice crimes become the most common category in the recidivism sample.

Table IV. Descriptive Statistical Evidence – Recidivism Duration Estimation

	16 years old	17 years old	18 years old	19 years old
Time elapsed before recidivism	347	358	405	489
Percentage increase		+3.17%	+13.12%	+20.74%
Number of individuals	6,818	5,051	7,446	4,421

Note: Time elapsed is in days before recidivism for a given individual.

Table IV shows a significant increase in the duration before recidivism between the ages of 17 and 18. However, the limited frequency data presents a challenge similar to that in Section 3.2.1. Specifically, although there is a significant rise in duration before recidivism between ages 16 and 17 compared to that between ages 17 and 18 (+3.17% versus +13.12%), it is difficult to establish whether the increase is discontinuous or linear.

4. Results

This study found that individuals who are treated more severely at the age of 18 have a decreased probability of recommitting a crime and a longer duration before recidivism after release. This suggests that harsher punishment can serve as a deterrent to criminal behaviour. Using the fact that individuals at 17 years old are treated differently from those at 18 years old in the criminal justice system, our results validate Becker's theory, which states that harsher punishment increases the perceived risk of reoffending, making it less likely that individuals will engage in criminal behaviour again. The longer time before recidivism also suggests that the increased perceived risk persists for some time after release from prison, indicating that harsher punishment can have a lasting deterrent effect. It is essential to note that the current study only focuses on the specific context of age-based differences in court sentencing.

One potential limitation to the interpretation of our results is the low frequency of the data used. Specifically, the study uses yearly observations to compare individuals who were treated differently based on their age. Using high-frequency data, such as weekly observations, would provide a more precise estimate of the effect of court sentencing on recidivism rates. For instance, 17 years and 51 weeks old could be used as a good counterfactual for 18 years and 1 week old, but this may not hold when comparing 17 years old to 18-year-old individuals. The use of low-frequency data could, therefore, introduce bias into the analysis and limit the generalizability of the findings. While acknowledging this limitation, it is important to note that the current study provides valuable insights into the role of court sentencing in deterring criminal behaviour, underscoring the potential benefits of harsher punishment. We will discuss the limitations in more detail in Section 5, but for now, let us focus only on the results of our study.

4.1 Logistic Model

The results of the logistic regression model indicate that receiving more severe treatment as an adult significantly decreases the likelihood of recidivism. The findings are presented in column (1) of Table V, which depicts the regression results without any control variables. The coefficient on the dichotomous adult variable of -2.7343 suggests that being treated as an adult has a negative impact on the probability of recidivism. Additionally, the calculated marginal effect of -0.2298 provides further evidence of this outcome, revealing that being treated as an adult decreases the likelihood of recidivism by approximately 23%.

Table V. Discontinuity in Probability of Recidivism

Variable	**(1)**	**(2)**	**(3)**	**(4)**
A. Estimates				
Adult	-2.7343*** (0.0453)	-2.7293*** (0.0453)	-2.7960*** (0.0463)	-2.7588*** (0.0464)
Sex		-0.1538*** (0.0443)	-0.1829*** (0.0445)	-0.2385*** (0.0448)
Length of case			0.0006*** (0.0000)	0.0005*** (0.0000)
Number of guilty				0.0420*** (0.0040)
Number of charges				0.0089*** (0.0020)
B. Marginal effects (Adult)				
No control variables	-0.2298*** (0.0039)			
With controls		-0.2293*** (0.0039)	-0.2341*** (0.0039)	-0.2274*** (0.0039)
R^2	0.1830	0.1833	0.1871	0.1981
N	**48,680**	**48,680**	**48,680**	**48,680**

Note: The table gives estimates of a logit model with recidivism as the dependent variable, taking the value 1 if a given criminal committed a second crime and 0 otherwise. The coefficient of discontinuity is estimated on the variable Adult. The variable sex takes the value 1 if individuals are male and 0 otherwise. The variable length of case represents the length of a given case in days. Number of guilty represents the number of individuals found guilty within a given case and number of charges represent the number of charges within a given case. Standard errors in parentheses. Statistical significance: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

The addition of various control variables to the base regression model results in a slight but consistent increase in the goodness of fit. The negative coefficient of the sex variable indicates that women are more likely to reoffend than men, which is consistent with the descriptive statistics presented in Table I that show a slightly higher proportion of women in the recidivism sample. The control variables in our model do not seem to have a significant impact on deterrence, indicating that the severity of sentencing is the primary driver of crime deterrence in our analysis. In other words, despite the inclusion of control variables, the calculated marginal effects consistently show a decrease of approximately 23% in the likelihood of recidivism when individuals are subjected to harsher treatment.

4.2 Duration Model

The results observed in the duration regression align with those derived from the logistic model. That is, being treated as an adult has a significant impact on reducing the event probability of recidivism. As seen in Table VI, using the coefficient of -0.2170 from column (1) and our base model equation 2.17, we compute a hazard ratio (or risk ratio) by dividing the hazard of recidivism for adults by the hazard of recidivism for youths $\left(\frac{h(a \geq 18)}{h(a < 18)}\right)$. Then, the results can be interpreted by subtracting 1 from the risk ratio and multiplying by 100 which leads to the estimated percent change in the hazard of recidivism for being treated in the adult criminal court system. For example, the hazard of recidivism for the regression from column (1) goes down by an estimated 19.5% for the adult group.

Table VI. Discontinuity in Hazard of Recidivism – Exponential Distribution

Variable	(1)	(2)	(3)	(4)
A. Estimates				
Adult	-0.2170***	-0.2173***	-0.1984***	-0.1904***
	(0.0130)	(0.0131)	(0.0131)	(0.0131)
Sex		0.0372*	0.0423**	0.0356*
		(0.0214)	(0.0214)	(0.0215)
Length of case			0.0038***	-0.0000***
			(0.0000)	(0.0000)
Number of guilty				0.0035*
				(0.0021)
Number of charges				0.0047***
				(0.0011)
B. Hazard (Adult)				
No control variables	-19.5033			
With controls		-19.5275	-17.9960	-17.3413
N	**23,634**	**23,634**	**23,634**	**23,634**

Note: The table gives estimates of a duration model with recidivism duration as the dependent variable (expressed in days) assuming an exponential distribution. The Standard errors in parentheses. Statistical significance: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Moreover, in column (5) of the regression results including all control variables, the coefficient on the dichotomous adult variable of -0.1904 indicates a negative effect on the probability of recidivism. Additionally, the calculated hazard of -17.3413 further supports this finding, suggesting that being treated as an adult reduces the probability of recidivism by approximately 17%. As a result, we can conclude that adults have a lower probability of recidivism during a given unit of time compared to youths.

5. Limits & Robustness

5.1 Limits on the Identification Strategy

To validate a regression discontinuity design, several robustness checks can be done. First, a placebo test could be performed by randomizing the assignment of adults and youths in a region near the cut-off of 18 years old to create a placebo group. For instance, we could simulate the age of majority to be 17 years old. If the regression results remained consistent for the placebo group, it would suggest that the originally estimated treatment effect for adults could be false. Second, alternative bandwidths could be tested since the estimated treatment effect for adults can be sensitive to the bandwidth. For instance, instead of using a grouping of the observations at the year level, we could test at the month, week, or day level. Third, covariate balance tests can be performed by examining the distribution of covariates across the cut-off point to ensure there are no significant differences between the adults and youths. Finally, sensitivity analyses can be conducted to examine the robustness of the results to alternative assumptions about the functional form of the model or other modelling choices.

However, a key limitation of this study is the use of low-frequency data, which may introduce bias into the analysis. Specifically, since the data is only available at the year level, it is difficult and/or ineffective to test the above-mentioned robustness tests. It is impossible to perform the placebo test as we observe cohorts only two years around the cut-off of majority (i.e., 16 to 19 years old). To observe a statistically significant number of criminals in our recidivism analysis, we observe all first-time criminals between the age of 12 and 15 and follow them until 19 years old. Simulating a false age of majority at 17 would only leave us with first-time criminals between 12 and 14 years old as a cohort to follow through time. As for the bandwidth test, since we do not have the choice to use bandwidth narrower than yearly, it goes without saying that it is simply impossible to change the bandwidth. The covariate balance tests also seem ineffective. In fact, individuals one year apart are very likely to be different on multiple factors that we are unable to observe such that these said factors could have a significant confounding effect on criminal behaviour. For instance, two cohorts of 17-year-old youth and 18-year-old adults could differ significantly in various social and psychological factors, such as finding jobs, quitting school, or moving out of their parents' house, which are not observed in the data and are therefore impossible to test for. That is, even if we were able to perform a balancing test on the covariates that we can observe, these unobserved variables could impact the results of the study and lead to inaccurate conclusions.

While higher frequency data would allow for testing the smoothness assumptions with accuracy and estimating the treatment effect, the use of low-frequency data limits the potential for such tests. Despite this limitation, we are still left with the last possible robustness check, which would be to do sensitivity analyses. That is, to ensure the reliability of the results we will test the sensitivity of the logistic and duration models to different assumptions and variations in the analysis. We will proceed by doing different regressions, but without mathematically modelling them as we did in Section 2.3. The goal is just to test our original analyses and to see if the result of the revised models still go in the same direction as our official ones.

5.2 Robustness Check of the Logistic model

A classical way to test the robustness of a logistic model is by using alternative models that are appropriate for binary outcomes, such as a linear probability model, for instance. That way, we could try to validate our result by comparing the sign of the coefficient on our explanatory variable. That is, we want to estimate the determinants of recidivism ($t = 1$) or no recidivism ($t = 0$) by regressing our variable of interest (dependent variable – dichotomic variable on whether criminals recommit a crime or not) on our variable of interest (independent variable – dichotomic variable on whether criminals are youths or adults) and our control variables. Because linear regression requires fewer assumptions about the distribution of the dependent variable and the error terms, it is more robust to deviations from these assumptions. This is the main benefit of using a linear regression as a robustness check of a logit regression. That way, a linear regression may be able to offer a complimentary assessment of the relationship between the being sentenced as an adult and recidivism.

Table VII. Linear Probability Regression – Validation of Logistic Model

Variable	(1)	(2)	(3)	(4)
A. Estimates				
Adult	-0.2172*** (0.0027)	-0.2167*** (0.0027)	-0.2193*** (0.0027)	-0.2123*** (0.0027)
Sex		-0.0149*** (0.0041)	-0.0167*** (0.0041)	-0.0207*** (0.0040)
Length of case			0.0000*** (0.0000)	0.0000*** (0.000)
Number of guilty				0.0050*** (0.0004)
Number of charges				0.0010*** (0.0002)
R^2	0.1188	0.1190	0.1207	0.1295
N	**48,680**	**48,680**	**48,680**	**48,680**

Note: The table gives estimates of a linear probability model with recidivism as the dependent variable. Standard errors in parentheses. Statistical significance: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

The results presented in Table VII are consistent with those from the logistic regression model shown in Table V. For instance, in column (4), the coefficient of -0.2123 suggests a 21% decrease in the linear probability of recidivism for the adult group. The R^2 value is actually smaller than the one from the logistic regression (0.1295 < 0.1981), indicating that assuming a linear relationship between treatment and the probability of recidivism may not be suitable. Nevertheless, the negative sign of the coefficient on the adult variable appears to support our earlier findings.

5.3 Robustness Check of the Duration model

As discussed in Section 2.3.2.2, we justify using an exponential survival distribution because we cannot affirm that the longer you spend outside of jail the higher the probability will be to (or not to) recommit crime, which is what a Weibull distribution would imply. However, we acknowledge that this assumption may not be realistic, as various factors such as aging, changes in living conditions, or changes in social networks can affect the risk of recidivism over time. To test the robustness of our results, we also implement a Weibull distribution in our survival regression analysis. This allows us to validate whether the impact of being treated more severely is consistent across different survival distributions.

Table VIII. Discontinuity in Hazard of Recidivism – Weibull Distribution

Variable	(1)	(2)	(3)	(4)
A. Estimates				
Adult	-0.1776*** (0.130)	-0.1778*** (0.0130)	-0.1611*** (0.0131)	-0.1547*** (0.0132)
Sex		0.0251 (0.0214)	0.0302 (0.0214)	0.0247 (0.0215)
Length of case			-0.0003*** (0.0000)	-0.0004 (0.0000)
Number of guilty				0.0034*** (0.0021)
Number of charges				0.0038*** (0.0011)
B. Hazard				
No controls	-16.2722			
With controls		-16.2890	-14.8793	-14.3328
N	**23,634**	**23,634**	**23,634**	**23,634**

Note: The table gives estimates of a duration model with recidivism duration as the dependent variable, assuming a Weibull distribution. Standard errors in parentheses. Statistical significance: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Column (5) of Table VIII displays the regression results, where the coefficient of the dichotomous adult variable of -0.1547 suggests that treating individuals as adults has a negative impact on the event probability of recidivism. Furthermore, the calculated hazard of -14.3328 provides additional evidence, indicating that being treated as an adult decreases the event probability of recidivism by around 14%. In light of this, assuming that the likelihood of recidivism could change over time, the instantaneous rate of arrest decreases in comparison to the exponential distribution (14% vs 17%).

Additionally, we can perform a linear regression using ordinary least squares (OLS) to determine if our findings align with the base model. The results of the simple linear regression are presented in Table IX. In this regression, we regress the duration in days (dependent variable) on our variable of interest (independent variable – a dichotomous variable indicating whether criminals are youths or adults) and our control variables. One advantage of using a linear regression as a robustness check for a duration model regression is that it is a relatively simple and interpretable model, which can supplement the interpretation derived from our duration regression results.

Table IX. Linear Regression – Validation of Duration Model

Variable	(1)	(2)	(3)	(4)
A. Estimates				
Adult	84.8153*** (6.2725)	84.7679*** (6.2727)	76.5065*** (6.3085)	73.7987*** (6.3170)
Sex		-9.5000 (10.3106)	-12.7711 (10.2920)	-10.2011 (10.2910)
Length of case			0.1579*** (0.0151)	0.1793*** (0.0155)
Number of guilty				-1.9017 (1.1588)
Number of charges				-1.8917*** (0.6036)
R^2	0.0076	0.0076	0.0121	0.0138
N	**23,736**	**23,736**	**23,736**	**23,736**

Note: The table gives estimates of a linear model with recidivism as the dependent variable. Standard errors in parentheses. Statistical significance: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

The linear regression results are interesting because they provide a different interpretation than previous regressions. In Column (1) of Table IX, we use a simple OLS regression to estimate the additional number of days before recidivism for those treated as adults. The coefficient of 73.7987 represents the increase in time before recidivism for adults. In other words, being treated more severely as an adult leads to approximately 74 days of additional time before recidivism. This finding further confirms the deterrence effect of severity on recidivism. The advantage of using a linear regression as a robustness check is that it provides a simple and easily interpretable model that complements the duration regression results.

6. Conclusion

The objective of this thesis was to evaluate the causal impact of punitive sentence severity on recidivism deterrence among offenders in Canada using a regression discontinuity design. This study adds to the existing literature in two significant ways. Firstly, it employs countrywide data set, containing more observations over a more recent and longer period. Secondly, it utilizes a model that considers the fact that recidivism is a time-to-event data, thus accounting for the timing of events and censored data.

We took advantage of the age cut-off at which individuals are systematically treated at the adult criminal justice system (i.e, more severely) to implement a regression discontinuity design. That is, the day individuals turn 18, they are subject to the criminal justice system, instead of the juvenile justice system.

In this context, the theoretical framework used followed Becker's theory of crime, which suggests that the severity of punishment plays a key role in deterring criminal behaviour. According to Becker, when individuals face the prospect of severe punishment for their criminal activities, they may be more likely to be deterred from engaging in criminal behaviour in the future. This is the case because the threat of punishment creates a sense of risk and uncertainty, making criminal behaviour less attractive. This deterrence effect, in turn, could lead to lower rates of recidivism. The general objective of this thesis was therefore to contribute to the growing body of research on the effectiveness of punitive measures by empirically validating (or not) Becker's theory, where the findings could have meaningful implications for the development of evidence-based policies aimed at reducing recidivism.

Precisely, the results suggest that individuals who are treated in the adult criminal justice system, tend to be deterred from recommitting crimes more than criminals that are treated in the juvenile justice system. In fact, the results from both the logistic and duration regression models suggest that being treated more severely (i.e., as an adult) has a significant impact on reducing the likelihood of recidivism. Being treated as an adult reduces the probability of recidivism by approximately 23%, according to the marginal effect calculated in the logistic regression model. The first robustness check (linear probability model) results confirm a decrease in the probability of recidivism by about 21% for the adult treatment group. Then, the hazard ratio computed from the duration regression model shows that the hazard of recidivism for adults goes down by an estimated 17%, and the second robustness check (hazard calculated from the duration regression assuming a Weibull distribution) indicates similarly that being treated as an adult reduces the event probability of recidivism by approximately 14%. The third robustness check (OLS regression of duration in days as the dependent variable) found that treating offenders more severely (i.e., as adults) increases the time before recidivism by about 74 days, indicating once more a deterrence effect of severity on recidivism. These consistent results across different models and specifications suggest that that an increase in severity of sentencing (i.e., being treated as an adult) is in fact a driver of crime deterrence. Overall, we can conclude that adults experience a significantly lower likelihood of recidivism compared to youths, and that therefore, being treated more harshly has a significant impact on reducing the probability of recidivism.

However, the use of low-frequency data is a major issue in this paper. In fact, our whole identification strategy is based on assumptions that the treatment group and control group are near perfect counterfactual. As discussed in Section 5, it is very likely that the smoothness assumption is therefore not respected, as we compare cohorts of individuals one year apart instead of weeks or even days. As said before, individuals one year apart might be very different on multiple factors that we were unable to observe (e.g., psychological changes, finishing high school, starting a new job, etc.) such that these said factors could have a significant confounding effect on criminal behaviour. If we compared the treatment and control groups one week apart in age, we could still capture the effect of the age cut-off while minimizing the potential for bias due to unobservable differences between the groups. By comparing the groups at ages close to the cut-off, we could ensure that they are similar in terms of observable characteristics such as age but are unlikely to differ in unobservable ways that might affect criminal behaviour. Despite this limitation, the current study provides valuable insights into the role of court sentencing in deterring criminal behaviour, highlighting the potential benefits of harsher punishment. Future research could build on these findings notably by using higher-frequency data to further explore the validity of observed impact of court sentencing on recidivism rates and advance our understanding of effective strategies for preventing criminal behaviour.

Bibliography

Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, *142*(2), 785-806.

Loeffler, C. E., & Grunwald, B. (2015). Processed as an adult: A regression discontinuity estimate of the crime effects of charging nontransfer juveniles as adults. *Journal of Research in Crime and Delinquency*, *52*(6), 890-922.

Lovett, N., & Xue, Y. (2018). Do greater sanctions deter youth crime? Evidence from a Regression Discontinuity Design, *SSRN Electronic Journal.*

Lee, D. S., & McCrary, J. (2005). Crime, punishment, and myopia. Unpublished Paper.

Mueller-Smith, M., Pyle, B., & Walker, C. (2023). Estimating the Impact of the Age of Criminal Majority: Decomposing Multiple Treatments in a Regression Discontinuity Framework. Working Paper.

Angrist, J. D., & Pischke, J. S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. *Princeton University Press*.

Eaton, S., Furness, H., & Brantigham, P. (2014). The Cost of Crime in Canada. *Fraser Institute.*

Gabor, T. (2016). Costs of crime and criminal justice responses. *Public Safety Canada, Sécurité Publique Canada.*

Buckley, J., & James, I. (1979). "Linear regression with censored data", *Biometrika*, *66*(3), 429-436.

Draper, N. R., & Smith, H. (1998). Applied regression analysis. *John Wiley & Sons*.

Febrianti, R., Widyaningsih, Y., & Soemartojo, S. (2021). The parameter estimation of logistic regression with maximum likelihood method and score function modification, *Journal of Physics, 1725*(1), 12-14.

Imbens, G. W., & Lemieux, T. (2008). "Regression discontinuity designs: A guide to practice", *Journal of econometrics*, *142*(2), 615-635.

Lee, D. S., & Lemieux, T. (2010). "Regression discontinuity designs in economics", *Journal of economic literature*, *48*(2), 281-355.

Becker, G. S. (1968). "Crime and punishment: An economic approach", *Journal of political economy*, *76*(2), 169-217.