MEASURING IMPUNITY

ALEXANDRE B. CUNHA

Abstract. In this paper I propose an estimator of the number of unpunished criminals. Its design takes into consideration the scarcity of data on the subject. Only two statistics are required to implement it, the number of punished offenders and the fraction of the offenses that led to punishment. Furthermore, a simple rule of three yields the estimate in question. The proposed estimator is unbiased if and only if the sample mean of the number of offenders per crime is an unbiased estimator of the population's corresponding parameter. A similar result holds for the efficiency and the minimum mean squared error properties. Among other findings, I provide evidence that during the period 1995-2018, a yearly average of 7,742 additional arrests would have taken place in the USA if the police had succeeded in clearing all murder and nonnegligent manslaughter cases. I also estimate that given their prison populations, between 106,383 and 594,588 murderers were not convicted in the USA and at least 61,830 killers avoided conviction in Mexico.

1. Introduction

Law enforcement is one of the most basic duties of a government. Whenever a criminal goes unpunished, the government has to some extent failed to fulfill this duty. Therefore, it seems natural to expect a considerable effort to be undertaken to assess the number of unpunished offenders. However, that does not seem to be case, since the literature in this respect is not extensive.

My goal in this paper is to contribute to fill this gap. I introduce a simple and parsimonious estimator of the number of unpunished offenders. Its design takes into account the scarcity of data on the subject. The estimator is a function of only two variables: the number of punished criminals and the ratio between punished and total (i.e., the sum of punished and unpunished) offenses. Having that information at hand, all that one has to do is to carry out a standard rule of three relating these two variables with the fraction of unpunished crimes. As illustrated by some applications, the estimator can be used in a large range of real-world cases.

For the purpose of this essay, an unpunished offender is anyone who participated in a crime and did not receive a penalty that was imposed by the law enforcement system on others who committed the same type of crime. A penalty can be an arrest that takes place when a crime is cleared by the police, a guilty verdict or any other (for instance, a fine). The estimator is sufficiently generic to be used with any definition of penalty.

Date: January 17, 2021.

Key words and phrases. crime, impunity, simple estimator, scarce data.

I gratefully acknowledge the financial support from the Brazilian Council of Science and Technology (CNPq). I also thank Eurilton Araújo and Luciane Carpena for their helpful comments. Any remaining errors are my sole responsibility.

The approach adopted in this paper is based on the fact that all occurrences of a type of crime in given location during a time interval (for instance, murders in the USA during 2014) constitute a population of crimes, while the subset of those crimes that led to punishment constitutes a sample. Therefore, the problem at hand consists in using the information provided by the sample to make inferences about the unsampled (i.e., unpunished) offenses.

As previously mentioned, data on impunity are scarce. For instance, a central variable in this respect is the number of offenders who participate in each crime. Therefore, ideally it would be known for each offense reported to the police how many individuals were punished as a consequence. Unfortunately, I did not locate any report or dataset containing this specific statistic. In spite of this obstacle, the estimator presented in this paper is sufficiently parsimonious to be applied with the available data and still yield some interesting conclusions.

I show in the paper that there is an equivalence between the problems of estimating the number of unpunished offenders and estimating the population mean number of offenders per crime. Indeed, there is a one-to-one relation between mean and impunity estimators; that is, for each mean estimator there exists exactly one impunity estimator and vice-versa. Moreover, an impunity estimator is unbiased if and only if its corresponding mean estimator has the same property. This conclusion holds regardless of the underlying probability distributions. The same type of necessary and sufficient results are true for the properties of efficiency (i.e., being unbiased and having the smallest variance in the class of unbiased estimators) and having the smallest mean squared error.

An attractive feature of the proposed impunity estimator is that its associated mean estimator is the well-known sample mean. Therefore, the estimator has the three properties mentioned in the previous paragraph if and only if the same is true for the sample mean. The fact that the scarcity of both data and previous studies on the matter is a major hindrance to make any inference about the probability distribution of the number of offenders per crime makes the result in question particularly convenient, since it establishes that the estimator presented in this paper is, in the sense of displaying these three properties or not, as good as a previously known and widely accepted estimator.

I evaluate the estimator's performance by means of some pseudo-out-of-sample predictions. The data source for this exercise is the FBI's (1996-2019) yearly report entitled *Crime in the United States*. This source allows me to build a dataset with yearly information on cleared crimes and arrests for six types of offenses (murder and nonnegligent manslaughter, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft) covering the period 1995-2018. For the sake of comparison, I also use two other estimators in these exercises. Both of them require much more information than the one introduced in this paper. I conclude that if the data are split in a manner that is consistent with its theoretical underpinnings, the proposed estimator can be accurate to the point of having all errors below 1.4% of the true value of the target variable and has better performance than the other two. On the other splits, it has a median performance when compared with the competitors.

I use the proposed estimator in five applications. In the first I estimate how many additional arrests would have happened if the police had cleared all crimes tallied in the aforementioned FBI report. The yearly averages of unrealized arrests are equal to 7,742 for murder and nonnegligent manslaughter, 347,661 for aggravated assault, 1,904,829 for burglary, 4,945,448 for larceny-theft, 770,617 for motor vehicle theft, and 304,152 for robbery. In the second, I apply data provided by the city of Chicago to conclude that during the period 2014-2017, on average there were 16,822 unrealized arrests per year for the crime of robbery. In the third application, I assess how many killers managed to avoid being convicted of murder and nonnegligent manslaughter in the USA. Due to the lack of data, I do not know the exact value of the ratio between punished and total crimes. For this reason, I cannot provide the precise value of the point estimate. However, I do find a minimum and a maximum value for the ratio in question and this allows me to estimate that, for those 186,521 prisoners serving time at the end of 2016 for murder and nonnegligent manslaughter, a minimum of 106,383 and a maximum of 594,588 killers managed to evade justice. In the fourth application, I carry out a similar exercise with Mexican data and show that for the 9,867 inmates guilty of homicide in 2018, an estimated minimum of 61,830 killers were not convicted. In the last application, I use the case of Canada to illustrate how to carry out an initial evaluation using some easily available statistics provided by the United Nations. This provisional assessment suggests that for the 1,805 people sentenced for homicide from 2003 to 2014 in that country, at most 4,361 killers were not punished.

[xxx IMPROVE xxx] Summing up, this paper makes a three-fold contribution. First, it presents an impunity estimator that can be used with the currently available data and may be useful even if more data become available in the future. Second, the paper also provides a blueprint of how to use any estimator of the mean number of perpetrators per crime to construct an estimator of the number of unpunished offenders. Third, the paper sheds light on the type of statistical information, beyond what is currently available, that is needed to advance the quantitative knowledge on impunity.

This paper is organized as follows. In Section 2 I briefly discuss some related works. In Section 3 I provide an overview of the problem. In Section 4 I present the estimator, discuss its statistical properties, and lay out the problem of estimating impunity in a ball-and-urn framework. In Section 5 I run a pseudo-out-of-sample evaluation. In Section 6 I carry out five applications of the estimator. I present my concluding remarks in Section 7.

2. Related literature

The literature on the problem of estimating the number of unpunished offenders seems to be very incipient. Indeed, I did not find a single essay aimed at this precise question. However, there are many works that discuss some closely related matters. I briefly discuss those whose conclusions are more relevant to this work.

I show in this paper that the problem of estimating impunity is equivalent to the one of estimating the mean of a distribution, in the sense that for each impunity estimator there is an associated mean estimator and vice-versa. Moreover, an impunity estimator has some basic statistical properties (such as unbiasedness) if and only if the same is true for its associated mean estimator. Furthermore, it is a simple task to use a mean estimator to construct an impunity estimator. In this respect, Catoni (2012), Devroye, Matthieu, Lugosi and Oliveira (2016) and Orenstein (2019) show that there are situations in which alternative estimators outperform the sample mean, which is the mean estimator underlying the impunity estimator introduced in this paper. Hence, provided that more data on impunity become available in the future, there is an open avenue to use the estimators studied by these authors to construct new impunity estimators.

The question addressed in this paper can be seen as the reverse of the problem of wrongful convictions. As Garrett (2020) points out, there is a vast literature on the latter issue. Among the several essays on the subject, the one by Norris, Weintraub, Acker, Redlich and Bonventre (2020) is particularly relevant for the present paper. They estimate that offenders who committed crimes for which innocent people were convicted will commit up to 41.7 thousand additional crimes per year in the USA. This finding suggests that a considerable number of transgressions can be a consequence of impunity.

Interestingly enough, the approach adopted here can potentially be used to estimate how many people were victims of judicial errors. Indeed, suppose that out of a set of crimes that led to a guilty verdict, a sample of them is audited for wrong verdicts. Having the result of this evaluation at hand, one can apply the estimator proposed in this paper to assess the number of people who were wrongfully convicted of those crimes that were not screened.

Several papers have studied the determinants of homicide clearing rates by the police. Some of them, such as Roberts (2007), Carter and Carter (2015) and Braga, Turchan and Barao (2019) focus on the United States or some of its localities; others, such as Roberts (2008) and Liem et al. (2019) carry out international comparisons. The findings of those papers have an interesting implication for this essay. The soundness of the estimator I introduce in this paper depends on whether or not the sample mean of the number of perpetrators per crime is a good estimator of the corresponding population mean. Furthermore, the available samples are composed of crimes that were cleared (or, besides being cleared, ended up leading to a guilty verdict). Not a single one of those papers lists the number of offenders responsible for the crime as a factor affecting the police's ability to clear a homicide. Moreover, none of the variables they identify as relevant (for instance, use of firearm in Roberts 2007 and investigative resources in Braga et al. 2019) correlates in an obvious fashion with the number of perpetrators. Hence, the existing literature on crime resolution does not suggest that the approach proposed this paper has a built-in bias.

3. An Overview of the Problem

In this section I present the overall structure of the problem at hand, establish some notation, obtain some simple formulas, etc. Moreover, I discuss the issue of data scarcity, which plays a central role in this paper.

3.1. **Counting criteria.** Before presenting the statistical framework, it is convenient to clarify a counting issue. A single person can participate in more than one crime (for instance, any serial murderer is responsible for multiple deaths). That brings up the following question: how should a criminal who was involved in multiple offenses be tallied? There are two obvious options: *single counting* and *double counting*. The former entails counting each offender just once regardless of the number of crimes she/he participated in, while the latter consists in counting each offender as many times as the number of crimes she/he committed.

There is a third possibility, which I call *mixed counting*. Two hypothetical examples illustrate its relevance. First, suppose that in a single year, a given police department uses single counting when tallying the arrests it carries out. However, it does not check whether a person that has just been detained had been arrested in previous years. As a consequence, despite the fact that the department is using single counting in each year, a single offender may be counted more than once across a time span of several years. Second, assume that, when clearing a burglary, a police officer arrests an offender who is soon released on bail. The next day, the officer finds out that the same person is responsible for two other burglaries and clears them by rearresting the criminal. Under single counting just one arrest should be tallied, while three should be computed under double counting. However, if the police department keeps track of the number of arrests, then it would add two to its records.

Ideally, researchers should be able to choose the counting criterion to be used. Nevertheless, usually that will not be possible. Instead, they will gather data that has been processed according to a given procedure and there will not be enough information to carry out a recount using another criterion. Hence, it is necessary to stick to the counting procedure adopted by those who tallied the raw data. Fortunately, as shown in Subsection 4.4, the estimator presented in this paper is consistent with all three tallying procedures. Of course, any resulting estimate will necessarily abide by the criterion used to count the criminals.

3.2. **Notation and basic relations.** The estimator discussed in this paper can be applied to study many different crimes, as well as several distinct types of punishments. For this reason, I often use generic terms like *crime* and *punishment* instead of specific ones such as murder, robbery, arrest, verdict, etc.

Let t_1 and t_2, $t_1 < t_2$, be two generic dates. Suppose that $N \geq 2$ crimes take place over the period from t_1 to t_2. The letter K denotes the total number of offenders[1] who participated in the N transgressions. Denote the ratio K/N by μ; that is, μ is the average number of people who took part in a crime. Assign an index $j \in \{1, 2, \ldots, N\}$ to each of those N offenses. Let S be the subset of $\{1, 2, \ldots, N\}$ such that if $j \in S$, then the corresponding crime led to a punishment.[2] Denote the number of elements of S by n and assume that $0 < n < N$. Now, let X_j be a random variable that is equal to the number of people that participated in crime j. As usual, x_j denotes a particular realization of X_j. Observe that

$$\sum_{j=1}^{N} X_j = K$$

and

$$E(X_j) = \mu, \tag{3.1}$$

where $E(\cdot)$ denotes the expected value.

My approach consists of treating the N crimes as a population and the n punished ones as a sample.[3] The task ahead consists of using the n sampled offenses to make

[1] Since the estimator presented in this paper can be used with any counting criterion, for the purposes of this section it is irrelevant which procedure was adopted to tally the criminals. With the exception of Subsection 4.4, the same is true for Section 4.

[2] This sanction can be applied to one or more persons; for instance, it is possible for two or more people to be convicted for a single murder.

[3] I assume there exist a population and a sample of crimes instead of a population and a sample of offenders. This procedure is consistent with the fact that the very first data of interest

inferences about the $N - n$ unsampled incidents. Denote the number of punished and unpunished criminals by, respectively, P and U.[4] Thus,

$$P + U = K,$$

$$\sum_{j \in S} X_j = P,$$

and

$$\sum_{j \notin S} X_j = U. \tag{3.2}$$

Since the sum in the last equality has $N - n$ terms, (3.1) implies that

$$E(U) = (N - n)\mu. \tag{3.3}$$

Finally, for future reference define the variable λ according to

$$\lambda = \frac{n}{N}. \tag{3.4}$$

Thus, λ is the fraction of the crimes for which at least one offender was punished.

3.3. **The matter of data scarcity.** The exposition contained in the previous subsection may suggest that after gathering information about the n solved crimes, I would know the values of t_1, t_2, N, n, and of each coordinate of the n-dimensional vector $(x_j : j \in S)$. However, as discussed next, usually that does not happen.[5]

By way of illustration, consider the yearly Crime in the United States report (FBI, 1996–2019). Take the crime of *murder and nonnegligent manslaughter* and define as the punishment the arrests made by the police when clearing a crime. Using the information provided there, it is possible to construct the table that follows.

are usually collected at the police level, and when that happens it is known that a crime took place, while the number of offenders responsible for it is either unknown or to be confirmed later.

[4] Both P and U are random variables.

[5] Actually, I am not aware of any dataset that is publicly available that contains all that information, either in a ready to use state or in such a way that it is possible to compute the values in question.

Table 1
Murder and nonnegligent manslaughter: basic data

year	(A) known offenses	(B) offenses cleared by arrest	(C) number of arrests	(D) arrests per cleared offense
1995	18,324	11,874	21,230	1.78794
1996	15,487	10,361	19,020	1.83573
1997	14,759	9,756	18,290	1.87474
1998	13,134	9,023	17,450	1.93395
1999	12,266	8,476	14,790	1.74493
2000	12,291	7,756	13,227	1.70539
2001	11,982	7,477	13,653	1.82600
2002	13,561	8,679	14,158	1.63129
2003	13,373	8,345	13,190	1.58059
2004	13,662	8,552	13,467	1.57472
2005	14,430	8,961	14,062	1.56924
2006	14,948	9,073	13,435	1.48077
2007	14,811	9,064	13,480	1.48720
2008	14,225	9,047	12,955	1.43197
2009	13,242	8,819	12,418	1.40810
2010	12,760	8,268	11,201	1.35474
2011	12,706	8,233	10,832	1.31568
2012	13,092	8,183	11,075	1.35342
2013	13,075	8,381	10,231	1.22074
2014	12,879	8,307	10,571	1.27254
2015	14,392	8,851	11,092	1.25319
2016	15,566	9,246	11,788	1.27493
2017	15,657	9,645	12,208	1.26573
2018	14,786	9,212	11,970	1.29939
total	335,408	213,589	325,793	1.52533

Sources: (A) and (C), FBI; (B) and (D), author's calculation using FBI data.

In this example, t_1 and t_2 are respectively equal to 1995 and 2018. The last line of the table contains some information concerning N, n, $\sum_{j \in S} x_j$, and $\bar{x}$. More specifically, $N = 335,408$, while the values 213,589, 325,793, and 1.52533 are *approximations* of n, $\sum_{j \in S} x_j$, and $\bar{x}$. I use the word *approximations* because some of these clearances are of crimes committed before 1995, while some of the crimes committed during the period 1995–2018 will be cleared in 2019 or later. A similar remark applies to the number of arrests, so the same is true for the average in the last column.[6] Concerning the vector $(x_j : j \in S)$, definitively it is not available. In summary, if one works with the Crime in the U.S., she/he will have at hand the values of t_1, t_2, N and approximations of n, $\sum_{j \in S} x_j$, and $\bar{x}$. It is not possible to recover the coordinates of the vector $(x_j : j \in S)$.

Clearly, the scarcity of data is a considerable hindrance. And in general this problem is even more serious than in the above example, because the Crime in the

[6]Since Table 1 covers a 24-year period, the approximations in question should be quite accurate for the whole time spam. Of course, the problem in question tends to be more severe with datasets that cover shorter periods.

U.S. is a very comprehensive document. In some applications even less data will be available. For instance, if one defines serving time in prison as the punishment, then she/he can take the number of incarcerated people as the value of $\sum_{j\in S} x_j$. However, it is unlikely that one will succeed in obtaining the values of t_1, t_2, and n. That is, suppose that one knows that 2,000 people are serving time for murder. Usually, it is not possible to establish either how many crimes were committed by those prisoners or the dates these offenses happened.

The last two paragraphs make clear that the design of an impunity estimator must take the scarcity of data into consideration. For this reason, the estimator I propose in this paper relies solely on two statistics: the sum $\sum_{j\in S} x_j$ (total number of criminals who were punished) and the ratio λ. In the context of the example I have just discussed, this entails using only the first three (from left to right) statistics in the last line of Table 1 to estimate how many additional arrests would have taken place if all 335,408 known offenses had been cleared.

4. Theory

This section is divided into four parts. In Subsection 4.1, I present the estimator; in Subsection 4.2, I analyze its statistical properties; in Subsection 4.3, I provide some formulas relating the estimator's accuracy to the sample and the population means; and in Subsection 4.4, I lay out the problem of estimating impunity in a setup of balls and urns.

4.1. **The estimator.** Before presenting the estimator, I have to discuss a theoretical issue. Standard statistical inference entails using a sample to make predictions about a population. For instance, the sample mean is often used as an estimator of the population mean (which is not a random variable). The problem studied in this paper is slightly different, since I use a sample to make predictions about the unsampled part of the population. More specifically, I want to make inferences about U. However, U is a random variable. Therefore, strictly speaking it is impossible to estimate U. What one can realistically expect to do is to estimate $E(U)$, and that is exactly what I do here. That being said, with small abuse of language I use expressions like "an estimator of U" and "to estimate U" throughout this essay.[7]

The proposed estimator, denoted by $\bar{U}$, is given by

$$\bar{U} = \frac{1-\lambda}{\lambda} P. \tag{4.1}$$

Therefore, all one needs to estimate U is the number P of criminals who were punished and the measure of the ratio λ of crimes that led to a penalty.[8] Given these two numbers, a simple rule of three yields the desired estimate. It should be clear that $\bar{U}$ simply imputes to the unsampled crimes a number of offenders that is proportional to the number of them found in the sampled incidents.

As usual, let $\bar{X}$ denote the sample mean $(1/n)\sum_{j\in S} X_j$. It is possible to express $\bar{U}$ as a function of N, n, and $\bar{X}$. To do that, combine the equality $P = n\bar{X}$ with

[7] I could have followed a different approach in this paper. Namely, I could have used the sample to estimate K (which is a population attribute) and from this result subtract $\sum_{j\in S} x_j$ to obtain an estimate of U. However, this procedure would not change the fact that U is a random variable. Given that the goal of the paper is to measure impunity, I decided to take the more direct approach of estimating U.

[8] As illustrated in Section 6, even those two statistics can be difficult to find.

(3.4) and (4.1) to conclude that

$$\bar{U} = (N - n)\bar{X}. \tag{4.2}$$

Hence, $\bar{U}$ estimates the number of unpunished criminals by simply multiplying the sample average number of criminals per offense by the number of unpunished crimes.

Of course, (4.1) and (4.2) are equivalent expressions. However, they play two very distinct roles in the remainder of this paper. As previously discussed, (4.1) takes into consideration the scarcity of data. I use it in the applications of Section 6. On the other hand, (4.2) spells out the relation between $\bar{U}$ and $\bar{X}$, making it particularly useful in the next subsection, where I study the statistical properties of $\bar{U}$.

4.2. **Statistical properties.** In this subsection I study under which conditions the proposed impunity estimator has three standard statistical properties. The first of them is unbiasedness. Recall that an estimator U' of U is unbiased if

$$E(U') = E(U), \tag{4.3}$$

while an estimator X' of the population mean μ is unbiased whenever

$$E(X') = \mu. \tag{4.4}$$

The second property is efficiency. Like many other authors, I say that an estimator is efficient if it is unbiased and has the minimum variance among all competing unbiased estimators. Formally, let $\mathrm{Var}(\cdot)$ denote the variance. An estimator U' of U is efficient if it is unbiased and $\mathrm{Var}(U') \leq \mathrm{Var}(U'')$ for every unbiased estimator U'' of U. An estimator X' of the population mean has to meet similar requirements to be efficient.

The last property consists of having the minimum mean squared error among all competing estimators. That is, an estimator U' has the minimum mean squared error among all estimators of U if $\mathrm{MSE}(U') \leq \mathrm{MSE}(U'')$, where $\mathrm{MSE}(\cdot)$ denotes the mean squared error and U'' is a generic estimator of U. For an estimator X' of the population mean to have this property, it has to meet an analogous condition.

I now outline the approach I adopt to investigate if $\bar{U}$ is unbiased, is efficient, and has the minimum mean squared error. Let U' and X' be, respectively, estimators of U and μ that satisfy the equality

$$U' = (N - n)X', \tag{4.5}$$

which clearly a is generalization of (4.2). I show that U' has any of these three properties if and only if X' has the same property. Since $\bar{U}$ and $\bar{X}$ satisfy (4.5), it follows that $\bar{U}$ has any of the three aforementioned properties if and only if $\bar{X}$ has the same property.

This line of attack has five benefits. The first three are substantiated in this paper, while the other two may be reaped if more data on impunity become available in the future. I discuss each of them below.

The first is to establish that, regardless of the underlying sample and distributions, the issue of whether or not the proposed impunity estimator has the aforementioned properties boils down to the problem of whether or not the sample mean, which is a well-known and widely used estimator, has those properties.

The second concerns the unbiasedness property. Recall that the sample mean of a simple random sample is an unbiased estimator of the population mean. Hence, if the n punished crimes constitute a simple random sample of the population of N crimes, then $\bar{X}$ is an unbiased estimator; hence, the same is true for $\bar{U}$. However, that is not all that can be said. Even if the n punished offenses do not constitute a simple random sample, $\bar{U}$ can still be unbiased. For that to happen, it is enough that the police apparatus and justice system that generate the sample do not introduce bias in $\bar{X}$.[9]

The third advantage pertains to the efficiency and the minimum mean squared error. Usually, assessing whether or not an estimator has these properties requires knowing the type of distribution of the data-generating process. However, the scarcity of both data and previous studies on impunity means that any such assumption would be groundless. Therefore, being able to state that $\bar{U}$ has those properties if and only if a popular estimator like the sample mean also has them turns out to be convenient to avoid the problem of missing information about the underlying distribution.

As previously mentioned, the exploitation of the last two benefits requires more data than currently available. Hence, for the remainder of this paragraph I assume that the condition in question is met. Concerning the fourth advantage, given that the results presented here are of the type "if and only if", it is enough to study the properties of the well-known estimator $\bar{X}$ to learn about the properties of the lesser-known estimator $\bar{U}$. Finally, the last benefit consists of providing a blueprint for creating an impunity estimator that satisfies a specific property. For instance, suppose that in some specific application it is essential to have an efficient estimator and $\bar{U}$ does not meet this requirement. A possible way to obtain an efficient impunity estimator consists of finding such an estimator for the population mean and then using equality (4.5) to obtain an estimator of U that has the desired property.

The following proposition and corollary formalize the results discussed above. The proofs are available in the Appendix.

Proposition 1. *Let U' and X' be, respectively, estimators of U and μ that satisfy (4.5). The following statements are true:*
(i) U' is unbiased if and only if X' is unbiased;
(ii) U' is efficient if and only if X' is efficient;
(iii) U' has the minimum mean squared error among all estimators of U if and only if X' has the minimum mean squared error among all estimators of μ.

Corollary 1. *The following statements concerning the estimators $\bar{U}$ and $\bar{X}$ are true:*
(i) $\bar{U}$ is unbiased if and only if $\bar{X}$ is unbiased;
(ii) $\bar{U}$ is efficient if and only if $\bar{X}$ is efficient;
(iii) $\bar{U}$ has the minimum mean squared error among all estimators of U if and only if $\bar{X}$ has the minimum mean squared error among all estimators of μ.

4.3. **The underlying means and the accuracy of $\bar{U}$.** The analysis carried in the previous subsection makes clear there is a close relation between the problems of estimating impunity and estimating a population mean. For the particular case of the estimator $\bar{U}$, this relation can be expressed in a very precise way. Indeed, it

[9]As discussed at the end of Section 2, the existing literature on homicide clearance does not suggest that such a bias exists.

is a simple task to show that

$$|\bar{U} - U| = N|\bar{X} - \mu|$$

and

$$\frac{|\bar{U} - U|}{K} = \frac{|\bar{X} - \mu|}{\mu}.$$

Hence, the absolute error of $\bar{U}$ is equal to the number of crimes times the absolute error of $\bar{X}$, while the absolute error of $\bar{U}$ as a fraction of the total number of offenders is equal to the absolute error of $\bar{X}$ as fraction of μ.

Given that $\bar{U}$ was designed to match the value of U, the ratio $\bar{U}/U$ provides a convenient metric of the estimator's accuracy. Now, observe that (3.2) and (4.2) imply that

$$\frac{\bar{U}}{U} = \frac{\bar{X}}{(N-n)^{-1} \sum\limits_{j \notin S} X_j}. \tag{4.6}$$

However, $(N-n)^{-1} \sum_{j \notin S} X_j$ is equal to the average number of offenders per unpunished crime. Therefore, the accuracy of $\bar{U}$ is exactly equal to the ratio between $\bar{X}$ and that other mean.

Equality (4.6) is particularly useful to interpret the results of Section 5, where I carry out some pseudo-out-of-sample predictions to evaluate the performance of $\bar{U}$.

4.4. **Balls and urns.** In this subsection I discuss the problem of estimating the number of unpunished criminals in a ball-and-urn setup. Initially I abstract from the counting criterion issue. After presenting the general structure of the model, I show that it can be applied regardless of the procedure used to tally the offenders.

4.4.1. *The framework.* Suppose that a large bag contains K balls. An agent, called *Nature*, randomly distributes those balls across N urns in such a way that each urn receives at least one ball (hence, $K \geq N$). Then, Nature selects $n < N$ of those urns, so that they are divided into two sets: the first of them contains the n selected urns, while the other contains the remaining $N - n$. Finally, Nature counts the total number of balls contained in the n urns of the first set. Now, suppose that a person, called *Researcher*, receives exactly the following three pieces of information from Nature: (i) there are two sets of urns containing balls; (ii) the total number of balls contained by the urns in the first set is equal to some value (denoted by P); (iii) the ratio of the number of urns in the first set to the number of urns in the second set is equal to some value (denoted by λ). Then, the Researcher is asked how many balls she/he expects to be contained by all urns in the second set.

An intuitive way to answer such a question is to apply a rule of three so that the number of balls in each set is proportional to the number of urns. However, this approach consists exactly of applying equality (4.1). This is due to the fact that the problem posed at the end of the previous paragraph is essentially identical to the one studied in the previous subsections. Indeed, it is enough to substitute the words *crime* and *offender* for, respectively, *urn* and *ball* in the previous paragraph to convert the problem at hand into the one of estimating impunity.

Given that similarity, it is possible to use this ball-and-urn example to shed additional light on the problem studied in this paper. The set of all crimes can be partitioned into two sets: the first is the set of punished crimes, while the second is the set of unpunished ones. Therefore, if ones knows the value of the ratio between the numbers of crimes in each set, then she/he can combine this piece of information with the number of offenders, as tallied by police and/or the judicial system, who participated in the crimes belonging to the first set and apply (4.1) to estimate the number criminals who participated in the crimes belonging to the second set. That is, the strategy of using the number of criminals who participated in the punished offenses to estimate the number of criminals responsible for the unpunished ones is equivalent to using the number of balls inside the sampled urns to estimate the content of their unsampled counterparts.

4.4.2. *Single, double, and mixed counting.* It is a relatively simple exercise to modify the above ball-and-urn example to explicitly deal with the counting issue. First, I show that the estimator can be used with either the single or the double tallying procedure. Then I show that it is also consistent with mixed counting.

Suppose, as before, that a large bag contains K balls. Every ball is marked with a number belonging to the set $\{1, 2, ..., K_0\}$. Each of these numbers is used at least once, implying that $K_0 \leq K$. An element of $\{1, 2, ..., K_0\}$ can be seen as a unique offender identification code, while the balls are used to describe her/his participation in the N crimes (i.e., urns). For instance, if there are balls marked with the number k in m urns, then the offender whose identification number is equal to k has participated in m crimes. As discussed below, these numbers can be used to avoid double counting the criminals.

Nature randomly distributes the marked balls across the N urns in such a way that each urn receives at least one ball. Furthermore, care is taken to ensure that no urn receives two balls that have the same identification number.[10] Then, Nature picks n urns so that they are divided, as before, into two disjoint sets.

At this point, Nature decides whether it wants to apply the single or the double counting criterion. If it selects the latter, then all it has to do is ignore the numbers in the balls and carry out exactly the steps described in Subsection 4.4.1. Of course, in that case the Researcher will be precisely in the position described in that subsection. If Nature wishes to adopt the single counting procedure, then it has to carry out an inspection of the urns. More specifically, it takes the following four steps (in the order they are presented):

(1) It takes an urn in the first set (i.e., the set with n urns), looks at its balls and writes down their numbers on a piece of paper.
(2) It takes a different urn in the first set, looks at its balls, compares their numbers with those written on the paper, removes from the urn those balls whose number were already recorded and adds to the paper the numbers of the balls that remained in the urn.
(3) It repeats the previous step until it finishes examining all urns in the first set.
(4) Still using the same piece of paper, it screens all the urns in the second set as described in steps 2 and 3.

[10] If a number had appeared twice in a single urn, then the same offender would have participated more than once in a single crime. Since that does not make sense, I rule out this possibility. I emphasize that it is a simple exercise to design a random allocation procedure that satisfies this constraint.

These four steps ensure that no urn will contain two balls with the same number after the inspection is concluded.[11]

After finishing the inspection, Nature tells the Researcher that (i) there are two sets of urns containing balls; (ii) the total number of balls contained by the urns in the first set is equal to some value (denoted by P_0); (iii) the ratio of the number of urns in the first set to the number of urns in the second set is equal to some value (still denoted by λ). Again, the Researcher is asked how many balls she/he expects to be contained by all urns in the second set. Of course, the Researcher faces exactly the same problem faced in Subsection 4.4.1. Therefore, she/he can answer according to the formula

$$\bar{U}_0 = \frac{1-\lambda}{\lambda} P_0,$$

where $\bar{U}_0$ is the conjecture on the number of balls contained in the urns of the second set. Except for the different notation, the last expression is identical to (4.1). Hence, the estimator proposed in this paper can be used regardless of whether the number of punished criminals was tallied using single or double counting. Needless to say, the resulting estimate will be consistent with the procedure used to tally the offenders.

I finish this part of the paper by discussing the mixed counting case. Suppose that before inspecting the urns, Nature divides them into groups (for instance, each group can correspond to a year). Then, the inspection is carried out in such a way to prevent double counting only inside each group. Clearly, the Researcher can still use expression (4.1) to assess how many balls are inside the urns in the second set.

5. Performance evaluation

Corollary 1 and Subsection 4.3 make clear that the performance of $\bar{U}$ is tied to the performance of the sample mean, which in turn is a well-known and widely used estimator. Despite that, it is worthwhile having an evaluation of the accuracy of $\bar{U}$ using real-world data. If not for other reasons, carrying out such an assessment is justified by the fact that in the first application of the next section I use the FBI's yearly Crime in the U.S. report to estimate the number of unrealized arrests (i.e., the additional number of arrests that would have taken place if the police had cleared all crimes) during the period 1995–2018 and, given the relevance of this exercise, it is important to have some understanding of how well $\bar{U}$ performs when applied to the FBI data. Therefore, in this section I use these statistics to run pseudo-out-of-sample predictions to assess how well $\bar{U}$ estimates U and compare its performance to those of some alternative estimators.[12]

[11]At this point, it is desirable to clarify two points. First, some urns may be empty after the inspection is finished. Regardless of that, the value of λ is not changed. After all, neither n nor N depends on whether offenders are single or double counted. Second, it is important that Nature examines initially the first set of urns and later the second. This ensures that if a given number shows up in a ball inside an urn in the first set and in another ball inside an urn in the second set, then the corresponding ball is removed from the urn in the second set while the other ball remains inside its urn. This is consistent with the idea that a criminal who committed two crimes but was punished for just one of them is tallied when assessing P and disregarded when evaluating U.

[12]The paper has a Data Appendix, which is available at http://www.alexbcunha.com/files/research/papers/paper18data.zip, containing all data used in this section, as well as a detailed description of their sources and instructions on how to replicate the results.

The Crime in the U.S. report provides detailed information about eight types of crimes: murder and non-negligent manslaughter, rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson. However, the definition of rape changed during the period in analysis (1995–2018). Furthermore, according to the FBI some law-enforcement agencies do not provide information on arson. Therefore, these two crimes are dropped from the analysis and I work with the remaining six.

I use the same approach for the six crimes. To illustrate it, next I discuss in detail how I carry out the pseudo-out-of-sample predictions for the crime category murder and non-negligent manslaughter. Consider Table 1. The idea is to partition the available 24 years into two subsamples and use one of them to predict the number of arrests in the other.

As pointed out by Clark and McCracken (2011) and Hansen and Timmermann (2012), there is no well-established rule or criterion on how to split the sample. However, the theoretical underpinnings of $\bar{U}$ provide a clear answer to this problem.[13] To understand why, assume for a while that the sample has to be split in half. Now, given the temporal structure of Table 1, it is natural to group the data in chronological order. Therefore, I partition the sample in such a way that the years are divided into the sets $\{1995, 1996, ..., 2006\}$ and $\{2007, 2008, ..., 2018\}$. One can think of this split as the one that would be carried out by a time-series researcher who opts to divide the sample into two groups of the same size. From now on I refer to these subsamples as, respectively, 1^{st} half and 2^{nd} half. Despite being intuitively appealing, there is a problem with this procedure. The estimator $\bar{U}$ is designed to use the information about the punished crimes and offenders and the unpunished crimes in a given time interval to estimate the number of unpunished offenders in the same time interval. On the other hand, a pseudo-out-of-sample prediction using the 1^{st} and 2^{nd} half partition requires the use of the data from the period 1995–2006 to predict the number of punished offenders in the period 2007–2018 and vice-versa. Therefore, the split in question is at odds with the conception of $\bar{U}$. In spite of that, this partition can be useful as a benchmark, so it is one of the splits I use to assess the performance of $\bar{U}$.

The discussion in the previous paragraph recalls the importance of carrying out the pseudo-prediction assessment in a manner that is consistent with the theoretical foundations of $\bar{U}$. This requires partitioning the sample in such a way that *data from a given period will be used to pseudo-predict the number of arrests in the same period*. However, there is no possible way of splitting the information in Table 1 to meet the condition spelled out in italics. The best that can be done is to divide the years in as in the sets $\{1995, 1997, ..., 2017\}$ and $\{1996, 1998, ..., 2018\}$.[14] Henceforth I refer to these subsamples as, respectively, odd years and even years. The odd/even years is the second split I use.

With the goal of observing how $\bar{U}$ performs in an adverse situation, I deliberately select a third partition in such a way that the its pseudo-predictions will be as off the mark as possible. As mentioned in Subsection 4.3, the ratio $\bar{U}/U$ is a convenient

[13] There is nothing new with the idea that the estimator being evaluated places constraints on the splitting options. For instance, there are 1,352,078 different manners of dividing the data in Table 1 so that each subsample contains 12 years. However, only one of those partitions can be used to carry out a time-series pseudo-forecast exercise.

[14] It should be clear that the split in question is the only way to partition the data so that no two or more consecutive years will be grouped together.

metric of the estimator's accuracy. Moreover, for the pseudo-prediction exercises to be carried out, equality (4.6) can be expressed as

$$\frac{\bar{U}}{U} = \frac{\bar{x}_i}{\bar{x}_o},$$

where $\bar{x}_i$ and $\bar{x}_o$ are, respectively, the in- and the out-of-sample averages. I use this fact to find the partition in which the ratio $\bar{x}_i/\bar{x}_o$ is the furthest from 1. The only additional constraint that I place is that the smallest subsample has to contain at least six years (i.e., 25% of the 24 available years).[15] These subsamples, which I call high $\bar{x}$ and low $\bar{x}$, constitute the last split I consider.

Summarizing, I split the sample in three different ways. In the first I do not take into consideration the principles underlying the design of $\bar{U}$ and simply partition the data in half in chronological fashion. In the second I try to match as close as possible the theoretical foundations of $\bar{U}$, so the years are grouped according whether they are odd or even. In the third, I purposefully select the subsamples so that the accuracy of $\bar{U}$ will be as low as possible, only requiring that each subsample must contain at least six years. The next table contains the data necessary to evaluate $\bar{U}$ for each of these splits

Table 2
Summary statistics by subsample for
murder and nonnegligent manslaughter

subsamples	P	n
1st half	185,972	108,333
2nd half	139,821	105,256
odd years	165,476	107,882
even years	160,317	105,707
high $\bar{x}$	117,660	64,723
low $\bar{x}$	208,133	148,866

Observe that P is the number of arrests (column C in Table 1), while n corresponds to the number of offenses cleared by arrest (column B in Table 1).

To illustrate how to carry out the pseudo-predictions, I use the odd year subsample to predict the number of arrests in the even years. I start by computing λ. Recall that the crimes in the odd years are supposed to be the punished ones, while those in the even years are treated as unpunished. Hence,

$$\lambda = \frac{107,882}{107,882 + 105,707} = 0.50509.$$

Combine this last result with (4.1) to conclude that

$$\bar{U} = \frac{1 - 0.50509}{0.50509} \times 165,476 = 162,140.$$

Since 160,317 arrests took place during the even years, the error of the pseudo-prediction was smaller than 1.2% of the actual value.

As exemplified above, instead of using the broader set of data contained in Table 1, $\bar{U}$ uses only the information available in Table 2. Ideally, I would compare the performance of $\bar{U}$ to those of alternative estimators that use the same amount of data. However, I am not aware of any estimator that satisfies that property. As a

[15]This constraint is irrelevant for the murder and nonnegligent manslaughter case. It binds for all other five offenses.

consequence, I must select estimators that require information beyond the contents of Table 2.

Although I do not require the alternative estimators to use only the information used by $\bar{U}$, I am still constrained by the data found in Table 1. Although this considerably restrains the search, I identified two alternatives. The first is the ordinary least square (OLS) estimator. I apply the data in Table 1 to run a standard linear regression, using the in-sample years, of the number of arrests against the number of offenses cleared by arrest. Then, I use the estimated equation and the number of offenses cleared by arrest in the out-of-the sample years to pseudo-predict the number of arrests for each of these years. In a last step, I add all the yearly pseudo-predictions to obtain an estimate of the number of arrests for the entire out-of-sample period. This last estimate is denoted by $\hat{U}$.

Concerning the second estimator, I use the fact that if I have an estimator X' (distinct from $\bar{X}$) of the population mean, then I can apply equality (4.5) to obtain an impunity estimator U'. Orenstein (2019, Subsection 2.5) discuss five alternative estimators of μ. Unfortunately, they all require more information than is available in Table 2. However, it is possible to implement one estimator that is very similar to the *median of means*.[16] As Orenstein explains, to compute the median of means one has to partition the sample into several groups of the same size, evaluate the sample mean of each group, and then take the median of those means. Now, observe that column D of Table 1 contains the mean number of arrests per crime for each year. Hence, if I treat the years as groups, the only thing that prevents me from using the data in column D to evaluate a median of means estimator of μ is the fact that the number of offenses cleared by arrest (column B) is not constant over the years (recall that the median of means estimator requires the groups to be of equal size). Nevertheless, I can compute a weighted median of the means. More specifically, I evaluate the in-sample median, weighted by the number of offenses cleared by arrest, of the yearly average arrests per cleared offense. I denote this estimator by $\tilde{X}$. Then, as prescribed by (4.5), I multiply $\tilde{X}$ by the total number of cleared crimes in the out-of-sample years to obtain an impunity estimator, which I call of $\tilde{U}$.

The next step consists of in presenting the metrics used to assess the performance of the estimators. Since the goal is to have an estimate of the value of U, probably the best metric for the performance of a generic estimator U' is the ratio U'/U. Thus, the fractions $\bar{U}/U$, $\hat{U}/U$, and $\tilde{U}/U$ constitute the primary measurement tool I consider. Of course, the closer a ratio is to 1, the better is the performance of the corresponding estimator.

The secondary metric I adopt is a normalized variant of the well-known root mean squared error (RMSE).[17] However, for this type of measurement tool to be meaningfully distinct from the ratio metric, it is necessary to have at least two pseudo-predictions for each split. And each of the estimators $\bar{U}$, $\hat{U}$, and $\tilde{U}$ provides just one pseudo-prediction per sample partition. Therefore, it is necessary to distribute their corresponding values across the out-of-sample years. Since $\hat{U}$ is equal to the sum over the out-of-sample period of yearly pseudo-predictions, I simply decompose it back into its constituent parts. Concerning $\bar{U}$, I use the fact

[16]The median of means estimator has been known at least since Nemirovsky and Yudin (1983).

[17]The reason for carrying out the normalization is that I compare the results of the pseudo-predictions across distinct splits.

that it is related to $\bar{X}$ as specified in (4.2). More specifically, I decompose $\bar{U}$ over the out-of-sample years according to the equality $\bar{U}_t = n_t\bar{X}$, where t is a generic out-of-sample year, $\bar{U}_t$ is the portion of $\bar{U}$ allocated to the corresponding year, and n_t is the number of offenses cleared by arrest during the year in question. In similar fashion, $\tilde{U}$ is decomposed as in the expression $\tilde{U}_t = n_t\tilde{X}$. I compute the normalized RMSE according to the formula

$$\frac{\sqrt{\frac{\sum_{t\in T_o}(U'_t-U_t)^2}{\#T_o}}}{\frac{\sum_{t\in T_o}U_t}{\#T_o}},$$

where t is an out-of-sample year, U' is any of the three estimators, U is the value to be estimated, T_o is the set of years outside the sample, and $\#T_o$ is its cardinality. The numerator in the above expression is equal to the RMSE of U', while the denominator is the yearly average of unpunished offenders.

The next six tables contain the results of the pseudo-predicion exercises. The first and second best results are highlighted, respectively, in bold and italics.

Table 3
Pseudo-predictions – murder and nonnegligent manslaughter

subsamples	ratio to U			normalized RMSE		
(in/out)	$\bar{U}$	$\hat{U}$	$\tilde{U}$	$\bar{U}$	$\hat{U}$	$\tilde{U}$
1^{st} half/2^{nd} half	*1.29229*	**1.28454**	1.31357	*0.29848*	**0.29186**	0.31939
2^{nd} half/1^{st} half	*0.77382*	**0.77492**	0.75693	*0.24491*	**0.24286**	0.26105
odd years/even years	*1.01137*	**1.00267**	1.03470	*0.14653*	**0.13993**	0.14966
even years/odd years	*0.98876*	**1.00087**	0.96539	*0.15291*	**0.14536**	0.15728
high $\bar{x}$/low $\bar{x}$	*1.30024*	**1.29750**	1.30604	*0.31483*	**0.31238**	0.32038
low $\bar{x}$/high $\bar{x}$	**0.76909**	*0.76263*	0.74522	**0.61094**	*0.62801*	0.67408

First and second best results in, respectively, bold and italics.

Table 4
Pseudo-predictions – aggravated assault

subsamples	ratio to U			normalized RMSE		
(in/out)	$\bar{U}$	$\hat{U}$	$\tilde{U}$	$\bar{U}$	$\hat{U}$	$\tilde{U}$
1^{st} half/2^{nd} half	*1.11016*	1.14498	**1.07923**	*0.11993*	0.15282	**0.09234**
2^{nd} half/1^{st} half	*0.90077*	**0.91103**	0.89468	*0.10997*	**0.10325**	0.11541
odd years/even years	**0.99732**	*0.99728*	1.00931	*0.06019*	**0.06011**	0.06033
even years/odd years	**1.00269**	1.02142	*1.00502*	**0.08086**	0.12968	*0.08097*
high $\bar{x}$/low $\bar{x}$	*1.13858*	1.14299	**1.11898**	*0.14114*	0.14545	**0.12195**
low $\bar{x}$/high $\bar{x}$	*0.87829*	0.87537	**0.87956**	*0.13244*	0.13539	**0.13126**

First and second best results in, respectively, bold and italics.

Table 5
Pseudo-predictions – burglary

subsamples	ratio to U			normalized RMSE		
(in/out)	$\bar{U}$	$\hat{U}$	$\tilde{U}$	$\bar{U}$	$\hat{U}$	$\tilde{U}$
1st half/2nd half	*1.09091*	1.10125	**1.07185**	*0.09964*	0.11069	**0.08296**
2nd half/1st half	*0.91666*	**0.93250**	0.91640	*0.09811*	**0.08594**	0.28960
odd years/even years	*0.99970*	0.99657	**1.00013**	0.06647	**0.06032**	*0.06644*
even years/odd years	**1.00030**	1.00981	*0.99671*	**0.06370**	0.06961	*0.06390*
high $\bar{x}$/low $\bar{x}$	*1.11883*	1.13272	**1.09827**	*0.12459*	0.13639	**0.10427**
low $\bar{x}$/high $\bar{x}$	*0.89379*	0.86440	**0.90037**	*0.11839*	0.14610	**0.11251**

First and second best results in, respectively, bold and italics.

Table 6
Pseudo-predictions – larceny-theft

subsamples	ratio to U			normalized RMSE		
(in/out)	$\bar{U}$	$\hat{U}$	$\tilde{U}$	$\bar{U}$	$\hat{U}$	$\tilde{U}$
1st half/2nd half	**1.08236**	*1.08399*	1.09499	**0.09419**	*0.09481*	0.10528
2nd half/1st half	*0.92391*	0.92040	**0.94369**	*0.09221*	0.09371	**0.07630**
odd years/even years	**1.00029**	0.99900	*0.99951*	*0.06471*	**0.06357**	0.06473
even years/odd years	**0.99971**	1.00571	*0.99499*	**0.06146**	0.06432	*0.06171*
high $\bar{x}$/low $\bar{x}$	1.12266	*1.12132*	**1.10563**	0.12300	*0.12180*	**0.10604**
low $\bar{x}$/high $\bar{x}$	0.89074	**0.89383**	*0.89179*	0.11892	**0.11546**	*0.11795*

First and second best results in, respectively, bold and italics.

Table 7
Pseudo-predictions – motor vehicle theft

subsamples	ratio to U			normalized RMSE		
(in/out)	$\bar{U}$	$\hat{U}$	$\tilde{U}$	$\bar{U}$	$\hat{U}$	$\tilde{U}$
1st half/2nd half	*1.18078*	1.23956	**1.17113**	*0.19504*	0.25096	**0.18614**
2nd half/1st half	0.84690	**0.90665**	*0.86836*	0.15817	**0.10160**	*0.13742*
odd years/even years	**0.98622**	0.97186	*1.02035*	0.08536	**0.04855**	*0.07949*
even years/odd years	**1.01398**	1.02901	*1.02319*	*0.09994*	**0.07871**	0.10004
high $\bar{x}$/low $\bar{x}$	*1.21086*	**1.09362**	1.22978	*0.23339*	**0.14375**	0.25089
low $\bar{x}$/high $\bar{x}$	*0.82586*	0.73053	**0.85833**	*0.19461*	0.30248	**0.16337**

First and second best results in, respectively, bold and italics.

Table 8
Pseudo-predictions – robbery

subsamples	ratio to U			normalized RMSE		
(in/out)	$\bar{U}$	$\hat{U}$	$\tilde{U}$	$\bar{U}$	$\hat{U}$	$\tilde{U}$
1st half/2nd half	1.09932	**1.09299**	*1.09828*	0.12866	**0.12184**	*0.12788*
2nd half/1st half	0.90966	*0.93565*	**0.94572**	0.11133	*0.09811*	**0.08314**
odd years/even years	*1.00349*	**0.99711**	1.01828	*0.07459*	**0.06429**	0.07591
even years/odd years	**0.99653**	*1.00666*	1.03028	*0.09615*	**0.09324**	0.09953
high $\bar{x}$/low $\bar{x}$	1.19007	**1.17716**	*1.18440*	0.19422	**0.18338**	*0.18862*
low $\bar{x}$/high $\bar{x}$	*0.84028*	0.80557	**0.84267**	*0.17114*	0.22111	**0.16884**

First and second best results in, respectively, bold and italics.

I identified seven major lessons from these pseudo-predictions exercises. I discuss each of them below.

First, all three estimators perform considerably better in the odd-even year split than in the other two partitions. To illustrate this point, take Table 1 and consider the estimator $\bar{U}$. Its ratios to U in the odd year/even year and even year/odd year pseudo-predictions are much closer to 1 than the corresponding ratios in the other four pseudo-predictions. This is true in all tables for all estimators. Furthermore, for any estimator considered, the normalized RMSE in the even-odd year split tends to be smaller than in the other two splits; the only two exceptions are $\hat{U}$ in Table 4 and $\tilde{U}$ in Table 8 (in both cases the normalized RMSE for even year/odd year is larger than its counterpart for 2nd half/1st half). The reason for that is simple. Recall that the third split is constructed in a manner that ensures that the ratio $\bar{x}_i/\bar{x}_o$ will be as far from 1 as possible. Since estimating impunity is equivalent to estimating the population mean μ, any impunity estimator tends to miss the target in that split. Concerning the poor performances in the first-second half partition, recall that column D in Table 1 corresponds to the yearly averages of the number of offenders per crime. That variable clearly has a decreasing trend. And a similar trend is present for the other five crimes too. Therefore, by splitting the data so that the first 12 years fall in one subsample and the remaining in the other subsample, one is unintentionally ensuring that the subsample means will differ from each other sufficiently enough for all estimators to perform poorly.

Second, as a general rule the discrepancies between $\bar{U}/U$, $\hat{U}/U$, and $\tilde{U}/U$ are small. That is, let U' and U'' be any two of the three estimators and consider the expression

$$100\left|\frac{U'}{U}-\frac{U''}{U}\right|.$$

For the odd-even partition, the highest value of that indicator is 4.9%; the remaining 11 are all smaller than 3.5%. For the first-second half partition, the highest three values are 6.8%, 6.6%, and 6.0%; all the other 9 are less than 3.7%. Concerning the high-low $\bar{x}$ split, the two highest values are 13.6% and 12.8%; the remaining 10 are smaller than 3.8%.

Third, all estimators perform very well in the odd-even partition. Under the ratio criterion the two worst results are those of $\tilde{U}$ in Table 7 (motor vehicle theft offense). However, even in that case both pseudo-predictions are off the target by less 3.5% of the true value of U. Furthermore, in the partition in question $\bar{U}$ pseudo-predicts U particularly well. Once more, the worst ratios are found in Table

7; and in both instances the error as a percentage of U is smaller than 1.4%. In Table 3 (murder and nonnegligent manslaughter), both errors of $\bar{U}$ are less than 1.2% of the target variable. For the remaining four offenses, the largest error of $\bar{U}$ corresponds to less than 0.4% of U.

Fourth, when the estimators are less accurate, they all err in the same direction. On the other hand, when they are close to the true value of U it may happen that one is below while another is above the target. Indeed, in each of the simulations for the first-second half and the high-low $\bar{x}$ splits the three ratios $\bar{U}/U$, $\hat{U}/U$, and $\tilde{U}/U$ are either all larger or all smaller than one, while this does not happen in the case of the odd-even partition.

Fifth, the ratio and normalized RMSE metrics yield the same estimator rankings in both the first-second half and the high-low $\bar{x}$ partitions and different ones in the odd-even split. This suggests that the two ranking criteria provide distinct results only when the estimators provide very accurate predictions.

Sixth, in the odd-even partition $\bar{U}$ clearly outperforms $\hat{U}$ and $\tilde{U}$ under the ratio metric. It has the best result in 8, the medium in 4, and the worst in 0 of the 12 exercises. The corresponding numbers for $\hat{U}$ are respectively, 3, 2, and 7, while 1, 6, and 5 are those of $\tilde{U}$. Under the normalized RMSE criterion, it seems that $\hat{U}$ has the best outcomes, followed by $\bar{U}$. This estimator has the best, the median and the worst result in 3, 7, and 2 exercises; concerning $\hat{U}$ and $\tilde{U}$, the corresponding numbers are 9, 0, and 3 for the former and 0, 5, and 7 for the latter.

Seventh, when compared to $\hat{U}$ and $\tilde{U}$, $\bar{U}$ tends to have an median performance in the first-second half and high-low $\bar{x}$ splits. In each of the two partitions, $\bar{U}$ has the best, the median and the worst pseudo-predictions in 1, 8, and 3 exercises. Concerning the other two estimators, in the first-second half case the corresponding numbers are 6, 2, and 4 for $\hat{U}$ and 5, 2, and 5 for $\tilde{U}$; in the high-low $\bar{x}$ partitions, the numbers are 4, 2, and 6 for $\hat{U}$ and 7, 2, and 3 for $\tilde{U}$.

I close this section with a brief assessment of the main implications of those seven lessons. Although being an obvious point, the first lesson provides a useful reminder that, as any other estimator, ideally $\bar{U}$ should be used only in a manner that is consistent with its theoretical foundations. In such a context, it seems to be able to provide very precise estimates of the number of unpunished offenders and to be more accurate than alternative estimators that require substantially more data to implement. If for any reason one uses, as in the first/second half and high/low $\bar{x}$ splits, $\bar{U}$ in a manner that is not fully in line with its conception, the second and the seventh lessons suggest that it is still capable of providing estimates that are on par with those of the competing and less parsimonious estimators.

6. Applications

In this section I use $\bar{U}$ in five different contexts. Besides contributing to illustrate the estimator's wide scope of applicability, some of those exercises have results that are relevant on their own.[18]

6.1. **Unrealized arrests in the USA.**

In this application I rely once more on the FBI's yearly *Crime in the U.S.* report. Again, I use the offense of murder and nonnegligent manslaughter as an example. The last line of Table 1 provides the

[18] As in the case of Section 5, this section is complemented by the Data Appendix available at http://www.alexbcunha.com/files/research/papers/paper18data.zip.

values of N (column A), n (column B), and P (column C) for the period 1995–2018. Thus, one can use expressions (3.4) and (4.1) to obtain an estimate of the unrealized arrests for the period in question. The same procedure can be applied to the other five offenses considered in Section 5. The results are available in Table 8. To save space, I use some abbreviations: MNM, AggAs, L theft, and MV theft stand for, respectively, murder and nonnegligent manslaughter, aggravated assault, larceny-theft, and motor vehicle theft. Since there are 24 years in the period, the last column on the right contains the yearly average of unrealized arrests.

Table 9
Unrealized arrests in the USA from 1995 to 2018

offense	N	n	P	λ	$\bar{U}$	$\bar{U}/24$
MNM	335,408	213,589	325,793	0.63680	185,814	7,742
AggAs	17,569,049	9,824,572	10,584,946	0.55920	8,343,862	347,661
burglary	42,619,807	5,565,511	6,866,472	0.13059	45,715,890	1,904,829
L theft	135,871,466	26,784,653	29,142,760	0.19713	118,690,760	4,945,448
MV theft	21,513,233	2,866,338	2,842,961	0.13324	18,494,816	770,617
robbery	8,204,886	2,236,390	2,735,167	0.27257	7,299,636	304,152

In each year about 7.7 thousand arrests for murder and nonnegligent manslaughter are not realized because the underlying crimes are not cleared; the corresponding number for the aggravated assault offense is 347.7 thousand.

6.2. **Unrealized robbery arrests in Chicago.** It is possible to use $\bar{U}$ with databases distinct from the FBI's Crime in the U.S. report. As an illustration, I carry out an application similar to the previous one using some statistics provided by the city of Chicago.

The Chicago Data Portal contains data on several topics, including public safety. It is possible to download a spreadsheet containing detailed information on each reported crime from 2001 up to seven days before the day of access.[19] The same spreadsheet also informs whether or not an incident led to an arrest. Thus, one can use the data in question to compute the total number N of offenses and the number n of offenses that led to punishment, as well as λ. The number of arrests P can be obtained at the Public Arrest Data portal of the Chicago Police Department, where records for each arrest that took place during the years in the period 2014–2017 are available.[20]

The two databases cover several types of offenses. Since the goal here is simply to provide an illustration of the applicability of $\bar{U}$, it is enough to pick just one of them to carry out the calculations. Hence, I select the crime of robbery (code 03). This encompasses several subcategories (for instance, *robbery armed: knife/cutting instrument* and *robbery armed: handgun*). For this specific offense, $N = 43,277$, $n = 3,883$, and $P = 6,632$. I apply (3.4) and (4.1) to conclude that $\lambda = 0.08972$ and $\bar{U} = 67,287$ for the entire period 2014–2017, while the yearly average of unrealized arrests is equal to 16,822.

[19] Mohler (2014) used exactly the same database.

[20] The information on whether or not a crime led to an arrest in the Chicago Data Portal is of the type *true* or *false*. Had the spreadsheet given the number of people arrested for each incident, then it would constitute an exception to the data scarcity problem discussed in this paper. Unfortunately, it is not possible to consolidate that spreadsheet with the arrest data provided by the Chicago Police Department to build a database containing information matching all variables discussed in Subsection 3.3.

6.3. **Unrealized murder and nonnegligent manslaughter convictions in the USA.** So far, arrests constitute the only type of punishment considered in this section. However, an arrested person will not necessarily be found guilty in a court of law. As a consequence, a crime that is cleared by an arrest may lead to no punishment (in the sense that nobody is convicted). For this reason, the lack of judicial convictions is a better metric to assesses impunity than the lack of arrests. Therefore, ideally one would use data on court verdicts when measuring the number of unpunished crimes.

Unfortunately, the implementation of the above prescription is far from simple. Indeed, that would require knowing, for those crimes reported to the police, the number of people who were found guilty of them in a court of law. As far as I am aware, there is no publicly available report that contains that information. What usually happens is a separation between police and court reports and/or datasets. At police level, one can find incident-based information about the number of offenses and the number arrests. On the other hand, most of the trial information is defendant-based; that is, for a given court, one can find information on how many people were tried, how many were convicted, and how many were acquitted. And there is no way to combine the information in the two types of reports so that one can carry out an exercise similar to those in the previous two applications, just substituting court convictions for police arrests. Despite this barrier, I show that $\bar{U}$ still allows one to provide an assessment of how many offenders manage to avoid conviction.

In the present application, I use the number of people who are serving time for murder and nonnegligent manslaughter as the measure of punished offenders P. Observe that it is not possible to identify the dates of the underlying crimes, the number of crimes each offender is serving time for, etc. In the notation of this paper, I have sound information about P but none about t_1, t_2, N, n, and the coordinates of $(x_j : j \in S)$. In spite of that, it is still possible to say something about the number of unpunished offenders. As equality (4.1) makes clear, the implementation of $\bar{U}$ requires knowledge of two variables, λ and P. I know the value of the latter variable. Concerning the former, my approach consists of obtaining a lower λ_l and an upper λ_h bound for λ and using those bounds to find an admissible range for the point estimate $\bar{U}$.[21] Indeed, the function $f(\theta, q) = [(1-\theta)/\theta]q$ is strictly decreasing in θ and $f(\lambda, P) = \bar{U}$. Thus,

$$\lambda_l \leq \lambda \leq \lambda_h \Rightarrow f(\lambda_h, P) \leq \bar{U} \leq f(\lambda_l, P).$$

The data provided in Bronson and Carson (2019) and Carson (2018) allow one to conclude that on December 31, 2016 there were 186,521 prisoners in American prisons serving time for murder and nonnegligent manslaughter. Therefore, I set $P = 186,521$. The value for the upper bound λ_h comes directly from Table 9. Indeed, according to information in that table, the police cleared 0.63680 of the crimes in question during the period 1995–2018. Since a crime that was not cleared cannot lead to a conviction, I set $\lambda_h = 0.63680$. Hence,

$$f(\lambda_h, P) = \frac{1 - 0.63680}{0.63680} \times 186,521 = 106,383.$$

Obtaining a meaningful lower bound for λ requires the use of sentencing statistics at both federal and state levels. Brown, Langan and Levin (1999), Durose and

[21] It is worth to point out that the range in question is not a confidence interval.

Langan (2003, 2004 and 2007), Durose, Levin and Langan (2001), Rosenmerkel, Durose, Farole (2009) contain information, consolidating all states, for the years of 1996, 1998, 2000, 2002, 2004, and 2006.[22] Federal data for those years is provided by the Bureau of Justice Statistics (1998, 2000, 2002, 2004, 2006, and 2009). I combine data from these sources with those of Table 1 to construct the table that follows.

Table 10
Cleared cases, arrests, and convictions

year	(A) number of cases	(B) number of cases cleared by arrest	(C) number of arrests	(D) number of people convicted	(E) maximum number of cases without a conviction
1996	15,487	10,361	19,020	11,694	7,326
1998	13,134	9,023	17,450	9,455	7,995
2000	12,291	7,756	13,227	8,883	4,344
2002	13,561	8,679	14,158	9,264	4,894
2004	13,662	8,552	13,467	8,590	4,877
2006	14,948	9,073	13,435	8,816	4,619
total	83,083	53,444	90,757	56,702	34,055

Sources: (A), (B), and (C), Table 1; (D), author's calculation using data from the above cited sources; (E), difference between (C) and (D).

Column E is obtained by subtracting column D from column C. As I explain next, this difference provides an upper bound for the number of cases without a conviction. To illustrate this point, consider the following question: given there were 34,055 more arrests than people convicted, what is the maximum number of cases in which nobody was convicted? One way to answer this question is to assign to those 34,055 "excess" arrests the largest possible number of cases. But that entails assigning to each of those arrests precisely one case. Thus, 34,055 is the answer to the question at hand.

Now, by subtracting 34,055 from the 53,444 cases cleared by arrests, I conclude that at least 19,839 cases led to at least one conviction. Therefore, I set the value of the lower bound λ_l according to

$$\lambda_l = \frac{19,839}{83,083} = 0.23879,$$

from which follows that

$$f(\lambda_l, P) = \frac{1 - 0.23879}{0.23879} \times 186,521 = 594,588.$$

Summing up, I estimate that for those $186,521$ offenders serving time in jail for murder and nonnegligent manslaughter at the end of 2016, there were at least $106,383$ and at most $594,588$ people who committed an equivalent crime but were not convicted.

I conclude this application by evaluating $\bar{U}$ over a discrete grid, contained in (λ_l, λ_h), for λ. This is done in the table that follows.

Table 11
Evaluation of $\bar{U}$ over a discrete grid for λ

[22]The information is provided in a report called *Felony Sentences in State Courts*. To my knowledge, the report is available only for the years considered in this paper.

λ	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60
$\bar{U}$	559,563	435,216	346,396	279,782	227,970	186,521	152,608	124,347

6.4. **Homicide impunity in Mexico.** According to the impunity index presented by Le Clercq and Rodríguez (2017), Mexico has the highest impunity level among the 21 countries from South, Central and North America included in that study. Homicides are often not punished. For instance, the data provided by Lecuona and Rodríguez (2019) indicate that in 2018 there were there a total of 32,715 intentional homicides in Mexico City and 28 of the 31 Mexican states. However, in the same year there were only 3,600 judicial convictions for this type of crime in those locations. Consequently, the Mexican case provides an interesting context to apply $\bar{U}$.

The approach here is similar to the one adopted in the previous application. Again, I use the number of people in prison to measure P. Concerning λ, it is not possible to pin down its value. However, I find a ceiling for that parameter and use this value to find a lower bound for $\bar{U}$.

To obtain the upper bound λ_h, I use the data provided by the organization Impunidad Cero at the website https://impunidata.org/datos-abiertos and few statistics from Lecuona and Rodríguez (2019) to construct the table that follows.

Table 12
Intentional homicide:
victims and guilty sentences in Mexico

year	2015	2016	2017	2018
victims	16,282	19,280	24,400	29,395
guilty sentences	N.A.	2,256	2,798	3,198

Obs.: data for Mexico City and 23 of the 31 states.

Let g denote the number of guilty sentences. For each punished crime, there must exist at least one such verdict. Therefore, $n \leq g$; hence, $\lambda \leq g/N$. Thus, I make $\lambda_h = g/N$. That being said, I set $g = 8,252$, which is equal to the total number of guilty sentences in the period 2016–2018. Concerning N, I use the total number of victims to measure it. However, both the numbers of victims and guilty sentences have increasing trends; since there is natural delay between a homicide and its trial, I would likely underestimate g/N if I used the number of victims in the period 2016–2018. For this reason, I set N equal to the total number of victims during 2015, 2016 and 2017. Hence, $N = 59,962$ and, as a consequence, $\lambda_h = 0,13762$.[23]

According to Mexico's National Institute of Statistics and Geography (INEGI), in 2018 there were 9,867 people serving time for homicide in Mexican prisons. Therefore, I set $P = 9,867$.

I can now provide a lower bound of $\bar{U}$. Since

$$f(\lambda_h, P) = \frac{1 - 0.13762}{0.13762} \times 9,867 = 61,830,$$

I estimate that for those $9,867$ people in Mexican prisons in 2018 for having committed homicide, there were at least $61,830$ other offenders who went unpunished.

6.5. **An initial assessment using readily available data.** After the needed data are at hand, the implementation of $\bar{U}$ is a simple and fast task. However, gathering the necessary statistics can be expensive. Hence, it would be desirable

[23]Had I used the period 2016–2018 to measure N, I would have concluded that $\lambda_h = 0,11293$.

to be able to carry out a preliminary sizing up of $\bar{U}$ to help decide whether or not to allocate resources to the process of data gathering. For this reason, in this example I show how to use some readily available information to carry out such an assessment.

The United Nations Office on Drugs and Crime (UNODC) provides a large range of statistics on crime, justice, and prisons at the website https://dataunodc.un.org/. It covers many nations. For instance, it contains the number of homicides in 93 different countries for 2018. At least for this specific crime, it provides enough information to perform an initial appraisal of the number of unpunished offenders. As an illustration, I consider the case of Canada. From 2003 to 2014, 1,805 individuals were convicted of intentional homicide. Thus, I set $P = 1,805$. The UNODC provides data allowing for a good tally of N. Unfortunately, the same is not true for n.[24] As a consequence, it is not possible to measure λ precisely. However, it is possible to obtain a lower bound λ_l for the variable in question. Indeed, a total of 6,166 people were brought into formal contact with the police (i.e., were suspected or even arrested) during the period in question. Let C_j be number of people contacted by the police while investigating homicide j. Since a punished offender must have been contacted by the police before being convicted,

$$\frac{P}{n} \leq \frac{\sum_{j \in S} C_j}{n}.$$

Assume that the C_j's have the same mean. Therefore,

$$\frac{E(P)}{n} \leq \frac{\sum_{j \in S} E(C_j)}{n} = \frac{\sum_{j=1}^{N} E(C_j)}{N} \Rightarrow \frac{E(P)}{E\left(\sum_{j=1}^{N} C_j\right)} \leq \frac{n}{N} = \lambda.$$

I now use the fact that $P = 1,805$ and $\sum_{j=1}^{N} C_j = 6,166$ to take the ratio $1,805/6,166$ as a tentative lower bound for λ. Thus, $\lambda_l = 0,29273$; as a consequence,

$$\bar{U} \leq \frac{1 - 0,29273}{0,29273} \times 1,805 = 4,361.$$

Therefore, this provisional assessment suggests that at most 4,361 offenders were not convicted of homicide in Canada during the period 2003–2014. On yearly terms, this corresponds to a maximum of 363 unpunished criminals.

7. Concluding remarks

In this paper I present an estimator of the number of people who committed a crime and were not punished by the law enforcement system. Its implementation requires knowledge of only two variables, the number of punished offenders and the fraction of the total offenses that led to punishment. A rule of three relating these two variables with the fraction of unpunished offenses yields the desired estimate. Despite the scarcity of data on the matter of impunity, the estimator's simplicity and parsimony makes it widely applicable.

I show that there is an equivalence between the problems of estimating the number of unpunished criminals and estimating the mean number of criminals that participated in each crime. More specifically, I establish that given a mean estimator

[24] According to the UNODC, during the period 2003–2014 there were 7,008 intentional homicides in Canada. So, one could set N equal to this last value. However, I did not find a way to combine the statistics provided by the UNODC to obtain even an estimated value of n.

that is unbiased and efficient (i.e., is unbiased and has the minimum variance among all unbiased estimators) and has the minimum mean squared error, it is a simple exercise to construct an impunity estimator that displays the same properties. The converse is also true. As a corollary of this equivalence, the proposed estimator has these three properties if and only if the same is true for the sample mean. Those results are true regardless of the underlying distributions.

To assess the performance of the estimator, I use yearly data provided by the FBI on six different types of crimes cleared by the police and the corresponding arrests during the period 1995-2018 to carry out pseudo out-of-sample predictions. For comparison purposes, I also use two other estimators, whose implementation requires much more data than mine, in these exercises. The results suggest that my estimator is very accurate (with errors smaller than 1.4% of the target value) and tends outperform the other estimators when the data are split in a way that is consistent with its theoretical foundations. When this condition is not met, it performs roughly on par with the other two.

I carry out five real-world applications of the estimator. Among other findings, I conclude that in the USA the fact that the police did not manage to clear all incidents of murder and nonnegligent manslaughter over the period 1995-2018 led to a yearly average of 7,742 unrealized arrests. The equivalent estimate for the crime of aggravated assault is equal to 347,661. I also estimate that for those 186,521 prisoners serving time at the end of 2016 for murder and nonnegligent manslaughter, between 106,383 and 594,588 killers avoid being convicted. Similarly, I show that for the 9,867 people guilty of homicide serving time in Mexican jails in 2018, there were at least 61,830 other killers who were not sentenced.

Appendix: proofs

Proof of Proposition 1. Consider Statement (i). Concerning its "if part", equalities (4.4) and (4.5) together imply that $E(U') = (N-n)\mu$. Combine this last result with (3.3) to conclude that (4.3) holds. For the "only if part", observe that (4.3) and (4.5) imply that $(N-n)E(X') = E(U)$. This last equality and (3.3) together imply that (4.4) is satisfied.

To establish the "if part" of (ii), assume that X' is efficient. Hence, X' is unbiased. Thus, Statement (i) implies that U' is also unbiased. Now, let U'' be any unbiased estimator of U, while X'' satisfies $X'' = U''/(N-n)$. As consequence of (i), X'' is an unbiased estimator of μ. Since X' is efficient, $\mathrm{Var}(X') \leq \mathrm{Var}(X'')$. Thus,

$$\mathrm{Var}(U') = (N-n)^2\mathrm{Var}(X') \leq (N-n)^2\mathrm{Var}(X'') = \mathrm{Var}(U'').$$

For the "only if part", suppose that U' is efficient. Therefore, U' is unbiased and, as a consequence of (i), so is X'. Now, let X'' be any unbiased estimator of μ and define U'' so that $U'' = (N-n)X''$. Statement (i) implies that U'' is an unbiased estimator of U. The fact that U' is efficient implies that $\mathrm{Var}(U') \leq \mathrm{Var}(U'')$, so that

$$\mathrm{Var}(X') = (N-n)^{-2}\mathrm{Var}(U') \leq (N-n)^{-2}\mathrm{Var}(U'') = \mathrm{Var}(X'').$$

The first step of the proof of (iii) consists of establishing that if the estimators X' and U' satisfy (4.5), then

$$\mathrm{MSE}(U') = (N-n)^2\mathrm{MSE}(X'). \tag{A.1}$$

Indeed, if (4.5) holds, then

$$U' - E(U) = (N - n)X' - E(U) \Rightarrow U' - E(U) = (N - n)(X' - \mu).$$

Square both sides of the last equality and then take their expected values to conclude that (A.1) is satisfied. It is now possible to show that the "if part" of (iii) is true. Suppose that X' has the minimum mean squared error among all estimators of μ and let U'' be any estimator of U. Define X'' so that $X'' = U''/(N - n)$. Since X' has the minimum mean squared error among all estimators of μ,

$$(N - n)^2 \mathrm{MSE}(X') \leq (N - n)^2 \mathrm{MSE}(X'').$$

Combined with (A.1), the last inequality implies that $\mathrm{MSE}(U') \leq \mathrm{MSE}(U'')$. Similar reasoning establishes the "only if part". □

Proof of Corollary 1. Since $\bar{U}$ and $\bar{X}$ satisfy (4.2), this result is a direct consequence of Proposition 1. □

References

Braga, Anthony A.; Turchan, Brandon and Barao, Lisa (2019). The Infuence of Investigative Resources on Homicide Clearances. *Journal of Quantitative Criminology* 35, 337–364.

Bronson, Jennifer and Carson, E. Ann (2019). *Prisoners in 2017*. Washington, Bureau of Justice Statistics.

Brown, Jodi M.; Levin, David J. and Langan, Patrick A. (1999). *Felony Sentences in State Courts, 1996*. Washington, Bureau of Justice Statistics.

Bureau of Justice Statistics (2009). *Federal Justice Statistics, 2006 - Statistical Tables*. Washington, Bureau of Justice Statistics.

Bureau of Justice Statistics (2006). *Compendium of Federal Justice Statistics, 2004*. Washington, Bureau of Justice Statistics.

Bureau of Justice Statistics (2004). *Compendium of Federal Justice Statistics, 2002*. Washington, Bureau of Justice Statistics.

Bureau of Justice Statistics (2002). *Compendium of Federal Justice Statistics, 2000*. Washington, Bureau of Justice Statistics.

Bureau of Justice Statistics (2000). *Compendium of Federal Justice Statistics, 1998*. Washington, Bureau of Justice Statistics.

Bureau of Justice Statistics (1998). *Compendium of Federal Justice Statistics, 1996*. Washington, Bureau of Justice Statistics.

Carter, David L. and Carter, Jeremy G. (2015). Effective Police Homicide Investigations: Evidence from Seven Cities with High Clearance Rates. *Homicide Studies* 20, 150–176.

Carson, E. Ann (2018). *Prisoners in 2016*. Washington, Bureau of Justice Statistics.

Catoni, Olivier (2012). Challenging the Empirical Mean and Empirical Variance: A Deviation Study. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 48, 1149–1185.

Clark, Todd E. and McCracken, Michael W. (2011). Testing for Unconditional Predictive Ability. In Clements, Michael P. and Hendry, David F. (eds.). *The Oxford Handbook of Economic Forecasting*. Oxford, Oxford University Press.

Devroye, Luc; Matthieu, Lerasle; Lugosi, Gabor and Oliveira, Roberto I. (2016). Sub-Gaussian Mean Estimators. *Annals of Statistics* 44, 2695–2725.

Durose, Matthew R. and Langan, Patrick A. (2007). *Felony Sentences in State Courts, 2004*. Washington, Bureau of Justice Statistics.

Durose, Matthew R. and Langan, Patrick A. (2004). *Felony Sentences in State Courts, 2002*. Washington, Bureau of Justice Statistics.

Durose, Matthew R. and Langan, Patrick A. (2003). *Felony Sentences in State Courts, 2000*. Washington, Bureau of Justice Statistics.

Durose, Matthew R.; Levin, David J. and Langan, Patrick A. (2001). *Felony Sentences in State Courts, 1998*. Washington, Bureau of Justice Statistics.

Federal Bureau of Investigation (1996–2019). *Crime in the United States, 1995–2018*. Washington, Federal Bureau of Investigation.

Garrett, Brandon L. (2020). Wrongful Convictions. *Annual Review of Criminology* 3, 245–259.

Hansen, Peter R. and Timmermann, Allan (2012). *Choice of Sample Split in Out-of-Sample Forecast Evaluation.* Unpublished manuscript.
Le Clercq, Juan Antonio and Rodríguez, Gerardo (coordinators, 2017). *Global Impunity Index 2017 (GII-2017).* San Andrés Cholula, UDLAP.
Lecuona, Guillermo R.Z. and Rodríguez, Paola G.J. (2019). *Impunidad en homicidio doloso en México: reporte 2019.* s.l., Impunidad Cero.
Liem, Marieke; Suonpää, Karoliina; Lehti, Martti; Kivivuori, Janne; Granath, Sven; Walser, Simone and Killias, Martin (2019). Homicide clearance in Western Europe. *European Journal of Criminology* 16, 81–101.
Mohler, George (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting* 30, 491–497.
Nemirovsky, A.S. and Yudin, D.B. (1983). *Problem Complexity and Method Efficiency in Optimization.* New York, Wiley.
Norris, Robert J.; Weintraub, Jennifer N.; Acker, James R.; Redlich, Allison D. and Bonventre, Catherine L. (2020). The Criminal Costs of Wrongful Convictions: Can We Reduce Crime by Protecting the Innocent? *Criminology & Public Policy* 19, 367–388.
Orenstein, Paulo N. (2019). *Topics in Robust Mean Estimation and Applications to Importance Sampling.* Ph.D. dissertation, Stanford University.
Rosenmerkel, Sean; Durose, Matthew R. and Farole, Jr., Donald (2009). *Felony Sentences in State Courts, 2006 – Statistical Tables.* Washington, Bureau of Justice Statistics.
Roberts, Aki (2008). Explaining Differences in Homicide Clearance Rates Between Japan and the United States. *Homicide Studies* 12, 136–145.
Roberts, Aki (2007). Predictors of Homicide Clearance by Arrest: An Event History Analysis of NIBRS Incidents. *Homicide Studies* 11, 82–93.

Federal University of Rio de Janeiro
E-mail address: research@alexbcunha.com
URL: http://www.alexbcunha.com