Criminal Discount Factors and Deterrence

Giovanni Mastrobuoni and David A. Rivers*

July 18, 2019

Abstract

The trade-off between the immediate returns from committing a crime and the future costs of punishment depends on an offender's time discounting. We exploit quasi-experimental variation in sentence length generated by a large collective pardon in Italy and provide non-parametric evidence on the extent of discounting from the raw data on recidivism and sentence length. Using a discrete-choice model of recidivism, we estimate an average annual discount factor of 0.74, although there is heterogeneity based on age, education, crime type, and nationality. Our estimates imply that the majority of deterrence is derived from the first few years in prison, which has important implications for policies related to crime and punishment.

*We thank Philip Cook, Greg DeAngelo, Zvi Eckstein, Alessandro Iaria, Igor Livshits, Lance Lochner, Salvador Navarro, and Daniel Nagin for helpful comments. We also thank seminar participants at the University of Essex, the University of Western Ontario, the University of Sienna, the University of Trier, IDC Herzliya, and Ben-Gurion University, and conference participants at the NBER Summer Institute on Law and Economics, the HCEO Conference on Social Interactions and Crime, the Royal Economic Society Annual Conference, the Summer Meetings of the Econometric Society, the Conference on Empirical Legal Studies in Europe, the International Association for Applied Econometrics Annual Conference, the Empirical Microeconomics Workshop in Banff, the Canadian Law and Economics Association Annual Meeting, the American Society of Criminology Annual Meeting, the Waterloo Economics Workshop, and the Society of Labor Economics for their comments. Giovanni Mastrobuoni is at the Collegio Carlo Alberto, University of Turin (ESOMAS) and University of Essex, E-mail: giovanni.mastrobuoni@carloalberto.org. David Rivers is at the University of Western Ontario, E-mail: drivers2@uwo.ca.

1 Introduction

Criminal activity involves an inherent trade-off between the immediate returns associated with committing the crime and the possibility of future punishments if caught. Since punishments are received in the future, the value that a potential criminal places on them depends on how that individual values future events. One implication is that individuals who more heavily discount the future will be more likely to engage in crime. Knowing the degree to which potential criminals discount the future is therefore a key component to developing a better understanding of what drives criminal behavior and for designing crime policy.[1]

This information is of substantial interest to policymakers and those involved in law enforcement. For example, in an effort to reform some of the mandatory minimum sentences enacted in the 1970s and 1980s, which contributed to prison population increases of 400 percent (Hunt (2015)), the US recently passed the First Step Act, a bipartisan prison and sentencing reform law which, among other things, reduces mandatory sentences for drug offenders and increases opportunities for early release. This reform would be consistent with policymakers realizing that longer sentences do not generate enough deterrence for drug offenders. Indeed, we find that drug offenders discount the future more heavily than other offenders.

Discounting has two important implications for deterrence. First, the more an individual discounts future events, the less deterrent power future punishments have. Second, discounting shapes the relative deterrent power of punishments received at different points in the future. For example, consider an offender who is choosing whether to use a *modus operandi* that might lead to an increased sentence (e.g., using a firearm, being violent, damaging property, etc.). As long as criminals discount the future, individuals considering crimes with longer baseline sentences will be less deterred by the same potential increase in sentence compared to those considering crimes with shorter sentences, as this additional time in prison would be served further into the future. The magnitude of this differential effect is increasing in the degree of discounting.

Discounting also has implications for the debate over the effect of severity of punishment on deterrence. In the original economic model of crime developed by Becker (1968), certainty and severity of punishment combine to form the expected cost of committing a crime. Since severity is typically associated with sentence length, it is commonly assumed that doubling sentence length leads to a doubling of the "costs" paid by offenders, and is therefore equivalent to a doubling of the probability of punishment (certainty). But this is only true if criminals do not discount the future, since increased sentence lengths are added on at the

[1] Discounting of the future has been found to be important in explaining individual behavior in a variety of settings, including savings/retirement (Cagetti (2003); Lawrance (1991); Warner and Pleeter (2001)); investment (Rust (1987)), risky behaviors/health (Harrison, Lau, and Rutström (2010); Viscusi and Moore (1989)), human capital investment (Keane and Wolpin (1997)), among many others.

end. Discounting of future consequences fundamentally changes this relationship, as it implies decreasing marginal deterrence of sentence length (severity). A consequence is that as discounting increases, the deterrent power of severity of punishment decreases relative to that of the certainty of punishment. The notion that swiftness and certainty of the penalty outmatch severity has been theorized at least as early as 250 years ago by Beccaria (1764).

In a survey of the recent evidence for deterrent effects of imprisonment, Durlauf and Nagin (2011) find that there is little evidence for strong marginal deterrent effects of increasing the severity of punishment, but considerable support in the literature for marginal deterrent effects of increasing the certainty of punishment (see e.g., Hawken and Kleiman (2009)). One explanation that they offer is low discount factors among criminals.

The idea that discounting is an important component of criminal behavior has been recognized in the literature at least since Ehrlich (1973). Cook (1980) highlights that with discounting and a constant per-period disutility of prison, increasing the severity of punishments will have a greater deterrent effect when the initial punishment is mild. Davis (1988) and Polinsky and Shavell (1999) develop formal models of intertemporal crime decisions that explicitly take into account the effect of discounting. Wilson and Herrnstein (1985) and Katz, Levitt, and Shustorovich (2003) suggest that one explanation for criminal behavior itself is low discount factors among those individuals that decide to commit crimes. However, despite prior recognition in the literature of the importance of discounting for criminal behavior (see also McCrary (2010)), there is very limited empirical evidence about the extent of discounting among criminals.

In this paper, we provide direct quantitative estimates of discount factors for criminals by taking advantage of a unique dataset related to a large-scale collective pardon passed in Italy at the end of July, 2006. In an attempt to reduce prison overcrowding, more than 20,000 inmates, corresponding to over one-third of the entire prison population, were released over a period of a few weeks. Inmates who had a residual sentence below three years at the time of the collective pardon were immediately released. A key condition of the pardon was that, if a released inmate was found guilty of a crime in the future, the pardoned sentence (or residual sentence) would automatically be added to the new sentence.

Our dataset consists of individual-level data on each released inmate. In addition to detailed information on the characteristics of each inmate, we observe the length of their original sentence that led to their pre-pardon incarceration, their residual sentence length, and whether or not they recidivate over a period of 17 months following the pardon.

Intuitively, identification of the discount factor is obtained by measuring the rate at which the marginal deterrent effect of imprisonment changes with sentence length. In other words, it is the shape of the relationship between the (dis)utility of imprisonment (which drives recidivism in the data) and sentence length that identifies the discount factor in our model. The presence of discounting implies that the marginal effect on offending of a longer sentence is decreasing in sentence length, as individuals discount the later years of a sentence more heavily. If the discount factor were equal to 0, then variation in the sentence would have no impact on the probability of recidivism, as individuals would place no value on future prison sentences. As the discount factor increases towards 1, each additional month of sentence has an increasing (negative) effect on the probability of recidivism. This leads to a convex decreasing relationship between offending (recidivism) and sentence length. The more convex the relationship between sentence length and the probability of recidivism, the larger the degree of discounting. We have illustrated this in Figure 1, by graphing the cumulative utility from prison as a function of sentence length for a discount factor of 0.74, which corresponds to our baseline estimate.

A key challenge for identification arises due to the fact that prison sentences are not randomly assigned. Judges impose sentences based on characteristics of the offender and the offender's criminal history, many of which will be unobserved to the researcher. This correlation between unobserved drivers of crime and observed sentence lengths generates an endogeneity problem, which has long been recognized in the deterrence literature (see Ehrlich (1973) for an early discussion and Levitt and Miles (2007) and Durlauf and Nagin (2011) for more recent summaries).

The natural experiment generated by the Italian pardon, however, provides a solution. Under the conditions of the pardon, inmates whose original incarceration happened at different times face different expected total sentences for the same future crime. Conditional on the length of the original sentence, the residual sentence only depends on the original date of entry into prison, which is plausibly exogenous and can be exploited to estimate the discount factor.[2] This source of exogenous variation in sentence length has been used previously by Drago, Galbiati, and Vertova (2009) to measure the average deterrent effect of imprisonment.[3]

An additional advantage of our analysis is that the policy we study is likely to be particularly salient. Our

[2]Conditioning on the original sentence is important, because individuals with longer original sentences are more likely to have longer residual sentences, as well as higher recidivism rates. This is because individuals that are deemed more likely to recidivate are more likely to be given longer original sentences, and mechanically, individuals with longer original sentences are more likely to have longer residual sentences at the time of the pardon.

[3]One issue that has been raised with this approach is that, *ceteris paribus*, inmates with a longer residual sentence served less of their original sentence in prison. To the extent that this additional time served in prison has a causal effect on recidivism, this might bias estimates based on variation in residual sentence length. In order to address this, we supplement Drago, Galbiati, and Vertova (2009)'s original data with the full history of previous incarceration spells for the sample of pardoned prisoners. This allows us to control separately for total time ever served in prison as well as the residual sentence. We discuss this more in Section 5 and provide empirical evidence that indicates that this is not biasing our estimates.

dataset consists entirely of individuals who have been previously incarcerated, and therefore are particularly likely to be aware of sanctioning rules (Kaplow (1990)). The California Assembly Study (1968) shows, not surprisingly, that knowledge of the maximum penalty for various FBI index type crimes is far better among incarcerated individuals (62 percent) compared to the rest of the population (25 percent).[4] Furthermore, the collective pardon led to a salient early release and applied simultaneously to a large number of inmates, which should have served to circulate knowledge and understanding of the policy.

Our estimates imply annual discount factors of 0.74 among previously convicted criminals. This degree of discounting suggests that while these individuals place a significantly lower value on the future, it is still the case that punishments, even those received several years in the future, entail non-negligible costs to a potential offender.

In order to highlight the importance of knowing the magnitude of discounting among criminals, it is useful to compare the effect of increasing sentence length for different values of the discount factor. Our estimated discount factor of 0.74 implies that doubling a 5-year sentence increases the disutility of prison by 22%, whereas doubling a 10-year sentence increases it by only 5%. If, instead, we employ a more traditional value (used in most economic contexts) of the discount factor of 0.95, these increases would be 77% and 60% respectively, suggesting a much larger potential role of sentence increases for deterrence, even at relatively long sentence levels. Our estimates, therefore, imply that increases in sentence length, either through mandatory minimums, sentence enhancements, or more severe sentencing, are unlikely to have much deterrent effect when the baseline sentence is already long.

Since our dataset contains detailed information on the characteristics of each inmate, we are able to examine differences in discount factors across many different dimensions of individual heterogeneity. Inmates with high education and those who commit crimes related to organizing prostitution exhibit discount factors that are close to 1 (0.99 and 0.92, respectively). The lowest discount factors are found for immigrants (0.66) and drug offenders (0.70). This heterogeneity in discount factors implies important differences in the deterrent effect of imprisonment across the population. For example, given our estimates, longer prison sentences are likely to generate more effective deterrence for offenders committing prostitution-related crimes compared to drug offenders.

There are a few studies in the literature that have taken advantage of survey questions designed to elicit time preferences or measures of impulsivity, and relate these to criminal behavior (see e.g., Nagin and Pogarsky (2004); Jolliffe and Farrington (2009); Åkerlund et al. (2016); and Mancino, Navarro, and Rivers (2016)). These papers generally find that individuals who are more present-oriented, or impulsive,

[4] See also Nagin (2013) and Lochner (2007) on the importance of criminals' perceptions in shaping crime.

are more likely to commit a crime. One drawback of these approaches is that it can be difficult to quantify the magnitude (as opposed to just the sign) of these relationships given the qualitative measures of time preference that are often used. There are also concerns regarding the measurement of time preference from laboratory/experimental studies (Coller and Williams (1999)). The paper most closely related to ours is Lee and McCrary (2017), who analyze recidivism rates for a group of released juvenile inmates using a regression discontinuity design around the age of majority (age 18) in the severity of sentencing. They find only a small decrease in offending at the age of 18, when expected sentence length increases, which is consistent with low discount factors.[5]

We also contribute to the non-experimental literature that has estimated discount factors for consumers (see the early studies by Friedman (1957) and Heckman (1976)) as well as for various other subpopulations (US military personnel, Warner and Pleeter (2001); US purchasers of durable and storable goods, Hausman (1979); Ching and Osborne (2019); homeowners, Giglio, Maggiori, and Stroebel (2015); Blevins et al. (2017); and US workers, Viscusi and Moore (1989)).[6] Our dataset contains only individuals who have been convicted of at least one crime, and therefore our results correspond to the subset of the population with a criminal record. Relative to the literature that estimates discount factors for the broader population, our results are on the low end of the range of estimates, particularly among those papers that use observational or non-experimental data. Thus, our estimates are consistent with high discounting serving as a determinant of crime.

The rest of the paper is organized as follows. In Section 2 we describe our dataset covering the collective pardon in Italy. Section 3 contains the description of our model and Section 4 details our identification strategy. In Section 5 we present our results and perform several robustness checks. In Section 6 we discuss the implications of our results for measuring deterrence. Section 7 concludes.

2 Data

Collective pardons are deeply rooted in Italian history. Since World War II there has been on average one pardon or amnesty every five years (Barbarino and Mastrobuoni (2014)), although in more recent years there have been only two, in 1990 and in 2006. Such pardons eliminate part of the inmates' sentences, typically two or three years, and inmates whose new net sentence drops below 0 are immediately released.

[5]We discuss their results and how they relate to our findings in more detail in Section 5.2.

[6]A series of papers also estimate discount factors using laboratory experiments (see e.g., Harrison, Lau, and Williams (2002); Harrison, Lau, and Rutström (2010); Meier and Sprenger (2010); Coller and Williams (1999)).

Given their wide reach, these pardons generate sudden releases of large numbers of inmates.[7]

On July 31, 2006 the Italian Parliament passed the last pardon, which was enacted shortly thereafter. We were given access to the incarceration spell of all prison inmates released on this occasion, including the exact dates of release and re-incarceration through December 2007. A key condition of the release is that if pardoned inmates are rearrested and convicted, the pardoned sentenced is added to the new sentence. Conditional on the initial sentence, this generates a plausibly exogenous source of variation in the expected severity of punishment.[8] This variation in expected future sentences allows us to measure how recidivism varies with sentencing and thus identify the discount factor. We also have information on inmates' nationality, age, education, and some other individual characteristics.

These are the same data used in Mastrobuoni and Pinotti (2015), and, compared to Drago, Galbiati, and Vertova (2009), allow us to follow the inmates an additional 10 months after release, for a total of 17 months. A large number of inmates were released (over 20,000 individuals representing more than one-third of the prison population).[9] In Table 1 we present summary statistics for the variables in our dataset. The average recidivism rate (within 17 months of release) is about 22%. Because our sample consists of individuals who have previously been convicted of a crime and been released from prison, the mean age in our sample (38) is higher than in many datasets on criminal offenders. One consequence of this is that these inmates are more likely to be familiar with the Italian criminal law and the rules governing their release. In other words, the deterrent effect of the law is likely to be salient. They are also likely to be more experienced offenders.

Overall the sample is quite uneducated. As shown in Table 1, about 27% of the individuals in our data have an education at or below primary school, and only 3% completed secondary school. About 62% of inmates are of Italian origin, 26% are married, 5% are female, and 15% were permanently employed prior to being imprisoned. 74% have a definitive sentence (i.e., exhausted all appeals).

Table 1 also provides a breakdown by crime type. Crime types are not mutually exclusive, as a crime can be both a property crime and a violent crime (e.g., robbery), and therefore the means of the indicator variables for each crime type do not sum to 1. Property crime is the most common type (58%), and prostitution and mafia-related crimes are the least common (2% each).

We supplement our main dataset with data on the incarceration histories of the inmates prior to the

[7] In the 2006 pardon a few crimes were excluded, including those related to the mafia, terrorism, kidnapping, and felony sex offenses, while other types of violent offenses, such as murder and armed robbery, were eligible.

[8] Technically, only individuals who receive a future sentence of at least two years are subject to serving their residual (pardoned) sentence. We discuss this in more detail in Section 5.4.

[9] As in Drago, Galbiati, and Vertova (2009), we drop individuals who were released while still awaiting trial, as we have no information about their sentence. In addition, we lose a small number of observations when matching the sample of released inmates used in Drago, Galbiati, and Vertova (2009) with a data extraction made by the Italian Prison Administration 17 months after the collective pardon.

pardon, obtained from an additional data request to the Italian Prison Administration. We were able to obtain the history of incarceration spells for over 90% of our sample. Of these, about two-thirds had incarceration spells prior to the one associated with the pardon. The summary statistics are provided in Table 1. On average these inmates had 3.67 prior incarcerations, with an average prior time served of just over 30 months.

We also collected auxiliary data on clearance rates and transition probabilities across crime types. Province-level data on the fraction of cleared crimes come from ISTAT (2005)'s criminal law statistics and are merged by province (there are 103 provinces in Italy) and crime type, with the individual-level recidivism data. The definition of a clearance in these data is that at least one suspect has been identified. As shown in Table 2, the province-level clearance rate is 23% on average, but with considerable variation both across location and crime type.

While we have information on the crime committed before the pardon, for inmates who are re-incarcerated all we observe for the new offense is the date of re-incarceration. Our main dataset also does not contain information on previous incarcerations that could be used to construct transition probabilities across crime types.[10] Instead, we take advantage of a dataset containing information on all inmates who served time in prison in one of two prisons located in Milan (Bollate and Opera) between 2001 and 2012. Using the entire history of offenses for these individuals (spanning from 1972-2012), we construct transition probabilities between crime types and merge these into our dataset. For inmates who commit a crime that falls into more than one category, we base the transition probabilities on the most serious crime, based on either the average or median sentence for each crime type. Table 3 shows that both measures generate similar transition rates. The large probabilities on the main diagonal show that criminals tend to recommit the same types of crimes.

3 Model

3.1 Baseline Model

Consider the decision problem of an individual i who has been pardoned and released from prison after committing a crime of type j. Given the conditions of the collective pardon described in Section 2, if this person commits another crime of type k and is caught, his/her total prison sentence, denoted ts_{ik}, will be equal to the new sentence, ns_{ik}, plus a residual sentence rs_i, where $ts_{ik} = ns_{ik} + rs_i$. Each individual decides whether or not to recidivate, and if so, which crime to commit. For now let us assume that $j = k$; the new

[10] Our supplemental data on incarceration histories does not contain a full set of crime-type variables, and thus we cannot use it to construct transition probabilities.

crime is the same as the original crime. We will relax this in Section 3.2.

Let $c_{ij} = 1$ denote that person i commits a crime of type j and $c_{ij} = 0$ otherwise. We normalize the utility associated with $c_{ij} = 0$ to be equal to 0. The net expected utility associated with committing a crime is given by:

$$V(c_{ij} = 1) = u\left(X_{ij}^{u}\right) + P_{jl}^{c} D\left(X_{ij}^{d}, ts_{ij}, \delta\left(X_{ij}^{\delta}\right)\right) + \varepsilon_{ij}. \tag{1}$$

The functions u and D capture the utility of committing a crime and the (dis)utility of being caught, sentenced to prison, and serving that time in prison.[11] X_{ij}^{u} denotes observed characteristics of the utility received from committing a crime that are related to the individual and the crime. X_{ij}^{d} denotes observed characteristics of the disutility of going to prison. X_{ij}^{δ} are observable characteristics that are related to differences in discount factors. P_{jl}^{c} is the probability of being caught and sentenced to prison for committing crime j in location l, and ε_{ij} captures unobserved drivers of the crime decision. δ is the discount factor, and the main object of interest here.

We assume that each year in prison leads to a constant flow disutility[12,13] of $d\left(X_{ij}^{d}\right)$ and discounting is exponential.[14] Given that $\sum_{t=0}^{T-1} \delta^{t} = \frac{1-\delta^{T}}{1-\delta}$, we have that[15]

$$D\left(X_{ij}^{d}, ts_{ij}, \delta\left(X_{ij}^{\delta}\right)\right) = d\left(X_{ij}^{d}\right)\left[\frac{1-\delta\left(X_{ij}^{\delta}\right)^{ts_{ij}}}{1-\delta\left(X_{ij}^{\delta}\right)}\right].$$

Letting $u\left(X_{ij}^{u}\right) = \alpha_0 + \alpha_1 X_{ij}^{u}$ and $d\left(X_{ij}^{d}\right) = \beta_0 + \beta_1 X_{ij}^{d}$, the expected (net) utility of crime can be written as:

$$V(c_{ij} = 1) = \alpha_0 + \alpha_1 X_{ij}^{u} + \left[\beta_0 + \beta_1 X_{ij}^{d}\right] P_{jl}^{c}\left[\frac{1-\delta\left(X_{ij}^{\delta}\right)^{ts_{ij}}}{1-\delta\left(X_{ij}^{\delta}\right)}\right] + \varepsilon_{ij}. \tag{2}$$

[11] In Appendix A, we show that the model for utility in equation (1) can be obtained from a more general dynamic utility model through a restriction on the option value of future crimes.

[12] This disutility includes the pain and suffering due to being imprisoned, the lack of freedom, as well as the opportunity cost of spending time in prison.

[13] This assumption is also employed in Lee and McCrary (2017). We are not aware of any estimates in the economics literature on the relationship between the length of a prison term and the per-period disutility of imprisonment. However, studies in criminology and psychiatry have found no evidence of a correlation between prison time and various measures of prison well-being, e.g., subjective quality of life, depression, anxiety, post-traumatic stress symptoms (see Bukstel and Kilmann (1980); Hochstetler, Murphy, and Simons (2004); Gullone, Jones, and Cummins (2000)).

While we believe that the assumption of a constant flow disutility of prison is a reasonable one, it may not even be necessary from a deterrence perspective. A more general interpretation of our results would be that of decreasing marginal returns to additional years of imprisonment. The policymaker cares about the deterrent effect of various sentence lengths when designing policies to reduce crime, and may not care if the source of diminishing marginal returns to imprisonment come from discounting or from a decreasing negative value placed on subsequent periods of punishment.

[14] In Section 5.4 we consider alternative forms of discounting such as hyperbolic discounting.

[15] The equation below implicitly assumes that the first month in prison is served immediately ($t = 0$). Our results are not sensitive to this.

The probability that a crime of type j is committed by individual i is then given by

$$\Pr(c_{ij}=1)=\Pr[V(c_{ij}=1)>0].$$

Equation (2) is the basis for our estimates in Section 5.

3.2 Allowing for the New Crime to be Different from the Old Crime

We now consider the case in which the new crime is not of the same type as the original crime (i.e., $j \neq k$). Let $c_{ijk}=1$ denote that individual i, who previously committed a crime of type j, commits a crime of type k after being released from prison. The net utility associated with this choice is

$$V\left(c_{ijk}=1\right)=\alpha_0+\alpha_1 X_{ik}^u+\left[\beta_0+\beta_1 X_{ik}^d\right] P_{kl}^c\left[\frac{1-\delta\left(X_{ik}^{\delta}\right)^{ts_{ik}}}{1-\delta\left(X_{ik}^{\delta}\right)}\right]+\varepsilon_{ik},$$

where $ts_{ik}=ns_{ik}+rs_i$. The probability of $c_{ijk}=1$ is then given by

$$\Pr\left(c_{ijk}=1\right)=\Pr\left[V\left(c_{ijk}\right)>0\right].$$

Recall that we only observe whether an individual is caught committing another crime and the day on which that occurred. We do not observe the type of the new crime type (e.g., violent, property, etc.). However, even though such information might be of use, it would still be unobserved for inmates who do not recidivate.

With a slight abuse of notation, we can denote the probability that individual i, who previously committed crime j, commits another crime of any type as an integral over all possible future crime choices k

$$\Pr(c_{ij}=1)=\sum_k \Pr\left(c_{ijk}=1\right)\Pr(k \mid j),$$

where $\Pr(k \mid j)$ are the probabilities of transitioning from crime type j to crime type k.

In the data, we directly observe the residual sentence rs_i and the original sentence os_{ij}. However, we do not observe the new sentence that an individual would receive for committing a new crime: ns_{ik}. The new sentence is likely to be a function of both the individual i and the crime type k. Therefore, we assume that the new sentences can be decomposed as $ns_{ik}=\theta_i as_k$, where as_k is the average sentence for a crime of type k, and θ_i is an individual-specific component of severity, reflecting violent attitudes of the offender, as well as any possible systematic sentencing differences based on characteristics of offenders that are observable to the judges. We can compute $as_k = \frac{1}{I_k}\sum_{i=1}^{I_k} os_{ik}$ from our data on original sentences, where I_k is the total number of people who committed crime k. We can then compute the individual component as $\theta_i = \frac{os_{ij}}{as_j}$. This implies that the new sentence for each possible crime choice k for a given individual i who previously committed crime j can be constructed as

$$ns_{ik} = \frac{os_{ij}}{as_j} as_k. \tag{3}$$

In the baseline model for which $j = k$, this implies that $ns_{ij} = os_{ij}$.

3.3 Relationship to Deterrence Models Without Discounting

The primary focus of our paper is estimating criminal discount factors. However, the model described above also has implications for the literature that attempts to measure the deterrent effect of imprisonment. In order to make this comparison clear, we will compare to the model used by Drago, Galbiati, and Vertova (2009), which uses a dataset very similar to ours to estimate the magnitude of deterrence, but the points we make apply more generally.

For simplicity, if we drop the characteristics X, and express $\left[\frac{1-\delta^{ts_{ij}}}{1-\delta}\right]$ as $\left[\frac{1-\delta^{os_{ij}}}{1-\delta} + \frac{\delta^{os_{ij}}(1-\delta^{rs_i})}{1-\delta}\right]$, we can re-write equation (2) as:

$$V(c_{ij}=1) = \alpha_0 + \beta_0 P^c_{jl}\left(\frac{1-\delta^{os_{ij}}}{1-\delta}\right) + \beta_0 P^c_{jl}\left(\frac{\delta^{os_{ij}}\left(1-\delta^{rs_i}\right)}{1-\delta}\right) + \varepsilon_{ij}. \tag{4}$$

This equation resembles the baseline equation that Drago, Galbiati, and Vertova (2009) estimate, which is given by the following:

$$y_i = \alpha + \beta_0 os_i + \beta_1 rs_i + \varepsilon_i, \tag{5}$$

where y_i is a binary variable for whether individual i commits another crime after being released and ε_i is assumed to follow a logistic distribution.[16]

There are two key differences between equation (4) and equation (5). First, the original sentence os and the residual sentence rs enter (4) non-linearly. Second, rs does not enter equation (4) separately from os. The importance of this is that, in Drago, Galbiati, and Vertova (2009), the effect of the residual sentence does not depend (directly) on the length of the original sentence.[17] Put differently, the first year of the

[16]Drago, Galbiati, and Vertova (2009) also estimate specifications which include characteristics of the individual and crime fixed effects.

[17]Because they estimate the model using a logit specification, this is not exactly true. Since the logit model is non-linear, the

sentence is assumed to have the same effect as the last year. This is a general feature of most models of deterrence in which the effect of additional years of imprisonment is not allowed to vary based on the length of the sentence onto which they are added.

In equation (4), however, the effect of *rs* depends directly on *os* (recall that $ns = os$ in our baseline setup). The intuition for this is straightforward. If an individual is caught committing another crime, the offender will receive a new sentence *ns*. In addition to this, the sentence will be extended by *rs*. In other words, after the *ns* months are served, the individual will need to serve an additional *rs* months. As long as $\delta < 1$, this distinction will be important. When $\delta < 1$, individuals discount the future, and the effect of an additional *rs* months in prison will depend on when that extra time is served, which depends on *ns*. The larger *ns* is, the further into the future the *rs* months are served, and the lower the effect on current behavior.

By L'Hopital's Rule, $\lim_{\delta \to 1} \frac{\delta^{os_i}(1-\delta^{rs_i})}{1-\delta} = rs_i$ and $\lim_{\delta \to 1} \frac{1-\delta^{os_i}}{1-\delta} = os_i$. Therefore, when $\delta = 1$, equation (4) becomes

$$V(c_{ij} = 1) \quad = \quad \alpha_0 + \beta_0 P^c_{jl}(os_i) + \beta_0 P^c_{jl}(rs_i) + \varepsilon_{ij},$$

which matches the form of the regression equation in Drago, Galbiati, and Vertova (2009). Thus their model can be seen as implicitly assuming that the discount factor is equal to 1. Since discount factors less than 1 generate marginal deterrent effects of imprisonment that are decreasing in sentence length, this implies that the estimate of deterrence in Drago, Galbiati, and Vertova (2009) is an average measure of deterrence across sentence lengths. This may help to explain one of their results. They find that for individuals with the longest original sentences, the effect of the residual sentence is negligible (pgs. 275-276). They attribute this to the fact that long sentences are given for serious crimes, and they suggest that more dangerous inmates (those that commit more serious crimes) are not deterred by prison. However, given the derivations above, an alternative explanation is that these individuals, if caught again, are likely to face a lengthy new sentence. The residual sentence will be served after that, and therefore far into the future. If individuals discount the future, then the residual sentence will have a low effect on utility, and thus it will not have much of an effect on decisions to recidivate.[18]

This analysis has important implications for studies of deterrence. To the extent that punishments

original sentence length will impact the effect of the residual sentence, but in the same way as any other covariate included in the regression. The original sentence length also has this effect in equation (4), but has an additional effect via its interaction with the residual sentence.

[18] More generally, our model predicts that when the discount factor is less than 1, the effect of the residual sentence on recidivism should be decreasing in the length of the original sentence, where recall that the original sentence is our proxy for the new sentence. Although they do not highlight this, this is exactly what Drago, Galbiati, and Vertova (2009) find, not just for those individuals with the longest sentences, but throughout the distribution of sentence length (see their Table 4).

are applied in the future (as is the case with imprisonment), failing to account for it might mask some important heterogeneity in deterrent effects. For example, sentence enhancements or other policies designed to lengthen sentences for very serious offenses (where baseline sentences tend to be long) will potentially have very small deterrent effects on crime.

4 Identification and Estimation

Under the assumption that ε_{ij} is independent of $\left(X_{ij}^{u}, X_{ij}^{d}, P_{jl}^{c}, ts_{ij}\right)$, one could estimate the model described by equation (2) via maximum likelihood. However, as stated earlier, the total sentence $ts_{ij} = ns_{ij} + rs_i$ is potentially correlated with the error ε_{ij}. There may be characteristics of the individual, which make that person more likely to commit crimes, that are unobserved to the econometrician, but observed by judges. As a result, individuals with higher ε_{ij}'s may also receive longer sentences, which generates a correlation between ε_{ij} and ns_{ij} (and os_{ij}). This also generates a correlation between ε_{ij} and rs_i, since individuals with longer original sentences are likely to have longer residual sentences as well. Fortunately, it is still possible to consistently estimate the parameters of the model. The key identifying assumption is that conditional on the original sentence received os_{ij}, the total sentence ts_{ij} is independent of the error. In Section 5.4 we discuss possible threats to identification and provide some additional empirical analyses to support this identifying assumption. This conditional independence assumption is based on the fact that the timing of the release is dictated by an (unanticipated) mass prison release. Therefore, the only source of endogeneity in the residual sentence rs_i is via the original sentence length. Once we condition on the length of the original sentence, the residual sentence is independent of the error.[19] Since $ns_{ik} = \frac{os_{ij}}{as_j} as_k$ (or $ns_{ij} = os_{ij}$ in the baseline model), controlling for the original sentence also controls for the endogeneity of the new sentence, as as_j and as_k do not vary by individual.

To summarize, the error in equation (2) is assumed to be correlated with the original sentence, but conditional on the original sentence, independent of the residual sentence. Therefore, as we discuss in more detail in Appendix B, we can control for the endogeneity by expressing ε as a function of the original sentence:

$$\varepsilon_{ij} = h\left(os_{ij}; \gamma\right) + u_{ij},$$

where h is a flexible function of os, and u is is assumed to be independent of os. By replacing this expression into equation (2), we have the following equation:

[19]See Drago, Galbiati, and Vertova (2009), which employs a similar assumption in order to estimate the deterrent effect of imprisonment, for a detailed justification of this assumption.

$$V\left(c_{ij}=1\right)=\alpha_0+\alpha_1 X_{ij}^u+\left[\beta_0+\beta_1 X_{ij}^d\right]P_{jl}^c\left[\frac{1-\delta\left(X_{ij}^\delta\right)^{ts_{ij}}}{1-\delta\left(X_{ij}^\delta\right)}\right]+h\left(os_{ij};\gamma\right)+u_{ij}, \tag{6}$$

where u is independent of $\left(X_{ij}^u, X_{ij}^d, P_{jl}^c, ts_{ij}, os_{ij}\right)$. The additional term $h\left(os_{ij};\gamma\right)$ controls for the dependence between the error and the total sentence ts_{ij} that would otherwise bias our estimates.[20]

5 Empirical Results

5.1 Non-Parametric Evidence

Before we present the estimates from our model, we first provide some non-parametric evidence from the raw data that illustrates the identification of the discount factor. Specifically, we compute the average recidivism rate for each of 40 quantiles of total sentence length, and plot them in the left panel of Figure 2.[21] The figure illustrates that as the total sentence length increases, the recidivism rate decreases, which is consistent with deterrence. Furthermore, the relationship is convex, suggesting that the marginal deterrent effect of imprisonment is decreasing, consistent with discounting. In the right panel, we add original sentence as a control in order to deal with the endogeneity of the total expected sentence.[22] To do this, we regress the recidivism indicator on a flexible function of the original sentence (a cubic polynomial in logs) and dummies for each of the quantiles of sentence length.[23] We then plot the relative coefficients on each of the dummies. There are two key changes in the figure. First, the overall magnitude of deterrence increases by a factor of almost four. Since sentence length is likely correlated with unobservable drivers of recidivism, we would expect the effect of increased sentence length to be biased towards zero. Once we add controls for original sentence to correct for this bias, we see much larger estimated deterrent effects. The second difference is that the predicted recidivism rates for each bin of sentence length are much more tightly concentrated along a curve that closely resembles Figure 1.

In the left panel of Figure 3, we add curves corresponding to both a fourth-degree polynomial and an exponential fit of the relationship, again controlling for original sentence. In the right panel we include several additional control variables in the regression. As both panels show, the exponential curve fits the data very well, consistent with exponential discounting. The implied annual discount factors of 0.77 and

[20] Essentially $h\left(os_{ij};\gamma\right)$ acts as a control function (Heckman and Robb (1985)).

[21] We trim the top and bottom 1% of total sentence length to make the figure easier to see.

[22] Conditional on the original sentence, the remaining variation in total sentence is due to the exogenous variation in residual sentence.

[23] Using splines and dummy variables based on the months of original sentence length led to very similar results.

0.81 are quite close to each other (suggesting that the inclusion or omission of controls is not driving our results) and to our baseline estimate of 0.74.

Overall, this descriptive evidence suggests non-trivial amounts of discounting that are apparent even in the raw data on recidivism and sentence length. The tight relationship between recidivism and sentence length foreshadows the high degree of statistical precision in our estimates, as well as the robustness of our results to alternative model specifications. In order to estimate the specific value of the discount factor, we now turn to the estimates from our model.

5.2 Baseline Results

We estimate our baseline specification described in equation (6) by maximum likelihood, using a logistic distribution for the error u, and we model $h(os)$ using a cubic polynomial in the logarithm of the original sentence.[24] In Table 4, we present estimates of the baseline model in equation (6). In our main dataset we do not observe the probability of being caught $\left(P^c_{jl}\right)$, and in our baseline estimates we assume that this is a constant. In Section 5.4 below, we bring in some additional external data on clearance rates to capture this probability, and find that our results are qualitatively unaffected. Intuitively, the reason why variation in clearance rates has little effect on our estimates of the discount factor is that clearance rates affect the disutility of all months of a future sentence in the same way, whereas the discount factor disproportionately affects later months, i.e., those served further into the future. Said another way, a decrease in clearance rates lowers marginal deterrence, but does not affect how marginal deterrence changes with sentence length, whereas a decrease in the discount factor lowers both marginal deterrence, and the rate of change in marginal deterrence. A similar argument explains why our estimates of the discount factor are not sensitive to the inclusion or omission of control variables in either the returns to committing a crime or in the per-period disutility of prison, as we discuss more below.

We start in column 1 with the simplest specification that includes no covariates. In other words, we investigate the effect of the residual sentence on recidivism, conditioning on only the original sentence length to control for the endogeneity of the residual sentence. Our estimates imply an annual discount factor of 0.71.[25] In columns 2-5 we add controls to the estimating equation to allow for the utility from crime commission and the disutility of imprisonment to vary across individuals. In column 2, we add controls for age, permanent employment status, gender, marital status, whether the inmate is Italian, whether the

[24] Model selection criteria show that this is the preferred specification for the original sentence. We also tried using splines and dummy variables based on the months of original sentence length, which led to very similar results.

[25] If δ is the monthly discount factor, then $\delta_A = \delta^{12}$ is the annual one.

original sentence is definitive (all appeals have been exhausted), and binary indicators for low education and high education. We define low education as having at most a primary education (corresponding to less than 5 years of education), and we define high education as an education at or beyond secondary school (corresponding to at least 13 years of education).

The first thing to note is that the estimated discount factor is relatively unchanged, having increased slightly to 0.74. Age is associated with a lower utility of crime commission, and a higher disutility of going to prison (although it is not precisely estimated in the first term). One explanation for this is that as people age, the expected benefit from committing crime is lower, perhaps due to a lower productivity of the individual in criminal activity, and serving time in prison is more unpleasant the older someone is.

Having a definitive sentence is associated with a higher disutility of prison. This is intuitive given that those without firm sentences face some chance of having their residual sentence commuted or shortened upon appeal. Being Italian is not associated with the utility from committing a crime, but is associated with a lower disutility of prison. This suggests that going to prison in Italy is worse for foreigners than for Italians. Since Italians have a higher chance of receiving alternative sentences such as home arrest and parole, this makes sense (e.g., according to ISTAT (2015), in 2013, foreigners represented 35 percent of the prison population but only 13 percent of the population receiving alternative sentences). Furthermore, in Italy, inmates are required by law to be imprisoned in the prison closest to their home, so that family can most easily visit. However, since 70 percent of foreign inmates are undocumented immigrants (Italian Ministry of Internal Affairs (2007)), by definition they have no official home, which means that they are likely to be imprisoned further from nearby friends and family. We also find that being married is associated with a lower utility of committing a crime and a lower disutility of going to prison. We find that gender, permanent employment prior to the original incarceration, and the level of education of the individual are not associated with either component of utility.

In columns 3 and 4 of Table 4, we estimate the model using alternative definitions of low education and high education. We first adjust low education to include only individuals with no completed education (column 3). In column 4, we re-define high education to be a high school degree or higher.[26] These changes have very little effect on the other estimates, including the discount factor. The point estimates on the indicators for low and high education do change a bit, but none of the estimates are statistically different from zero. The differences between these education measures do matter below, however, when we interact them with the discount factors in Section 5.3.

In column 5 we add dummies for the various crime categories, with the idea that the returns to commit-

[26] Compared to our original measure of high education we exclude individuals who went to vocational school.

ting crimes and/or the opportunity cost of going to prison might vary by crime type. The other coefficients are largely unchanged, and as was the case in the previous results, the estimated discount factor is relatively unaffected, with a point estimate of 0.70.

Overall, the results in Table 4 indicate that the (annual) discount factor among our sample of previous criminal offenders is between 0.70 and 0.74. In line with the non-parametric evidence shown in the second panel of Figure 2, the discount factor is precisely estimated. The 95% confidence interval for our baseline result in column 2 of is [0.65,0.83], and thus we can easily reject the hypothesis that the discount factor is either 0 or 1.[27]

In order to illustrate how discounting affects the relationship between utility and imprisonment, in Figure 4, we have graphed the present discounted flow disutility of prison as a function of sentence length, for various values of the discount factor. We normalize the disutility to 100 in period 0. Under no discounting (discount factor of 1), the relationship would be a straight horizontal line at 100. As the figure illustrates, even small changes in the discount factor lead to large relative changes in utility, and therefore behavior, which is what allows us to obtain precise estimates of the discount factor.

Our estimates imply that the disutility of an additional year of imprisonment is 74% of the disutility of the previous year. This indicates a fairly large degree of discounting by criminal offenders, but not so much that they act myopically and ignore any future consequences of their actions. Relative to the first year in prison, the marginal discounted annual disutility of prison drops to 30%, 7%, and 1% after 5, 10, and 15 years, respectively. The effect diminishes over time at a reasonably fast rate, but not so fast that the difference between short- and medium-term sentences becomes meaningless from a deterrence perspective. In other words, there is scope for a deterrent effect of imprisonment.

For a more traditional discount factor of 0.95, the discounted marginal disutility of imprisonment falls to 50% of its original value after 15 years. If we compare to smaller discount factors—at values of 0.10 and 0.30—the effect drops to below 1% after only 3 and 5 years, respectively, reflecting negligible marginal deterrent effects of imprisonment for sentences of more than a few years. At a discount factor of 0.74, there is some drop-off in the deterrent effect, but it does not vanish immediately. If discount factors were much smaller, then this would imply almost no deterrent effect of imprisonment beyond the first year or two.

Estimates of the individual discount factor for the general population, based both on lab experiments (e.g., Coller and Williams (1999); Harrison, Lau, and Williams (2002); Andersen et al. (2014)) and observational studies (e.g., Viscusi and Moore (1989); Warner and Pleeter (2001); Cagetti (2003)), vary considerably in the literature, ranging from values close to 0 to close to 1. However, the majority of the estimates

[27] Standard errors, and therefore confidence intervals, for the annual discount factors can be computed via the delta method.

range from discount factors of 0.7 to close to 1. Our estimates are on the low end of those found in the literature, particularly compared to those from other observational or non-experimental studies. This is consistent with the idea that those individuals who decide to commit crimes have lower discount factors than people in the population at large, and suggests that low future time preference could be an important driver of criminal behavior.

The only other paper that we are aware of that attempts to elicit individual discount factors for criminals is Lee and McCrary (2017). Using detailed panel data on arrests in Florida, they exploit the discontinuity in the severity of punishment at the age of majority (18). Despite large increases in the average severity of punishment at the age of 18, they find very small drops in the probability of arrest. Using a dynamic extension to Becker's (1968) model, they map these estimates into plausible values of the discount factor, and find that the youth they study are essentially myopic. Specifically, they find that their point estimates are inconsistent with annual discount factors larger than 0.022, although they note that they cannot statistically rule out larger ones.

There are several factors that likely contribute to the difference between our estimates and those in Lee and McCrary (2017). First, it is likely that the effect of the policy in our sample is much more salient. Individuals in our dataset are older, with an average age of 38, compared to 18 in Lee and McCrary (2017), and therefore likely to be more familiar with the criminal justice system, particularly the adult system.[28] Individuals in our sample are also quite well-informed about the future consequences of their actions, both in terms of the legal implications (a longer sentence) and the costs associated with being imprisoned. Tied to their early prison release, which should be apparent to them, is the caveat of having to serve the residual sentence if they are convicted again. Furthermore, the massive scale of the pardon and the homogeneity of the policy across individuals (with the important exception of the length of the residual sentence) should have served to increase the flow of information regarding the pardon across prisoners.

Second, given the age differences in our samples, it is possible that younger people act more myopically (particularly those who decide to commit criminal acts), perhaps due to problems with self-control or impulsivity (Wilson and Herrnstein (1985); Jolliffe and Farrington (2009)). It is difficult for us to compare directly, as no one in our sample is less than 19 years of age, and fewer than 1% are below the age of 22.

[28]Using NLSY data for the US, Hjalmarsson (2009a) finds that youth perceptions of the change in the sanctions at the age of majority (age 18) are smaller than the true changes observed in the data.

5.3 Heterogeneity in the Discount Factor

In order to investigate whether there are any important differences in discounting behavior across individuals we also allow for discount factors to vary across individuals based on observables. We first allow the discount factor to vary based on characteristics of the inmates, such as gender, age, education, nationality, etc. Second, we allow the discount factor to vary based on the category of the original crime committed (e.g., violent, property, drugs). This analysis not only provides a description of how discount factors vary (or do not vary) across individuals, but also provides a check on our results, as it facilitates comparison to other results in the literature regarding individual discount factors.

We first estimate the model allowing for the discount factors to depend on individual characteristics, by interacting the discount factor δ with indicator variables for marital status, gender, low education, high education, employment status, as well as age and measures of nationality. These results are presented in Tables 5 and 6. In columns 1 and 2 of Table 5, the point estimates suggest higher annual discount factors for inmates who are married and for women (6 percentage points (pp) and 9 pp, respectively). Although the differences are not precisely estimated, our results with respect to gender are consistent with the literature (see e.g., Cagetti (2003); Coller and Williams (1999)). We are not aware of any results in the literature regarding a systematic relationship between discount factors and marital status. We find little evidence of differences based on whether or not an individual was permanently employed at the time of imprisonment prior to the pardon.

We find evidence that discount factors are decreasing with age. In columns 4 and 5 we report results in which we allow the discount factor to depend on age in a log-linear[29] and quadratic fashion, respectively.[30] In both cases we find evidence that discount factors decrease in age at a decreasing rate. For the log-linear specification the discount factor decreases from 0.78 at age 26 (the 10th-percentile of the age distribution) to 0.68 at age 51 (the 90th percentile). With the quadratic model, the discount factor drops from 0.76 at age 26 to 0.68 at age 51. Overall there seems to be evidence of a somewhat steep decline in discount factors early on, followed by a flattening out. There is no consensus in the literature on the relationship between time preference and age, with some papers finding that discount factors decrease with age (e.g., Meier and Sprenger (2010)) and others finding the opposite (e.g., Warner and Pleeter (2001)). One potential explanation for our finding of a negative relationship in our dataset is selection. To the extent that low discount factors are an important driver of crime, individuals who are still committing crimes later in life are likely those who have the lowest discount factors.

[29] Specifically we interact the discount factor with the natural log of the age of the offender minus 18 years.

[30] We also tried interacting the discount factor with dummies based on different age bins, and found similar results.

For education, we consider two definitions of both low and high education, following the definitions used previously. For low education (columns 6 and 7), the differences in discount factors for both definitions are small (less than 2%) and statistically insignificant, although the signs are what we would expect (lower educational attainment is associated with lower discount factors). One likely explanation is that the average education level of individuals in our dataset is quite low, so distinguishing between low and very low education levels has negligible effects on the estimates of the discount factor.

In contrast to the measures of low education, we do find sizable differences for high education. The results in columns 8 and 9 indicate that having a higher education is associated with significantly larger discount factors. For individuals with an education of at least 13 years, the estimated discount factor is 0.89 compared to the rest of the released inmates at 0.73. When we exclude from the high education measure those that attended vocational school, the difference is even larger, with an estimated discount factor of 0.99 for those with high education, a difference of 26 percentage points. While less than 2% of the sample have this high of a level of education, the difference is very precisely estimated, suggesting that particularly high levels of education are strongly related to discount factors. Given the nature of our data, we are unable to determine whether education actually increases discount factors (Becker and Mulligan (1997); Lochner and Moretti (2004)) or whether people with high discount factors are more likely to invest in education. Regardless of the explanation, consistent with the literature, we find strong evidence of a positive relationship between (high) education and discount factors (Viscusi and Moore (1989); Warner and Pleeter (2001); Cagetti (2003)).

The last set of individual characteristics that we analyze are based on nationality. Recall that we observe the country of origin for each individual. We first estimate a specification in which we allow the discount factor to depend on whether the person is Italian or not. The results in column 1 of Table 6 indicate a discount factor of 0.78 for Italians compared to 0.66 for non-natives. This result could be driven by a number of factors. First, it could be the case that the discount rates of the populations from which the immigrants originate are lower than in Italy. It could also be the case that those individuals who both immigrate and commit crimes have particularly low discount factors.

Salience might be important here as well. Immigrants may be less familiar with the Italian justice system and they may not speak Italian, both of which could lead to a failure to fully recognize the consequences of their future criminal activities, and therefore lower estimated discount factors in the data. Regardless of the explanation, one implication of our estimates is that foreigners are less sensitive to increases in sentence length.

In addition to examining differences between Italian natives and immigrants, we examine how differences in criminal discount factors vary among immigrants across country of origin. A recent literature has emerged documenting an empirical cross-country relationship between measures of time preference and many aggregate behaviors and outcomes such as national income, investment in human and physical capital, and savings rates (e.g., Galor and Özak (2016) and Dohmen et al. (2015)). In a recent attempt to determine the cause of these underlying differences in time preference, Chen (2013) examines differences across countries based on the extent to which their languages distinguish between current and future events, which linguists refer to as future time reference (FTR). His hypothesis is that speaking in a way that places a larger distinction between current and future events causes people to value the future less. In support of this, Chen finds a negative relationship between FTR and future-oriented behaviors such as savings and various health-related activities. In other words, in countries in which the language places a stronger distinction between the current and future, individuals exhibit less forward-looking behavior on average.[31]

In order to investigate the presence of differences in discounting across immigrants based on country of origin, we merged in the data on FTR from Chen's paper. We matched the primary languages spoken in each country[32] to the country of origin for each individual. Each language is coded by linguists as having either strong FTR or weak FTR. If all major languages spoken had either strong or weak references, we classified the country accordingly. For example, the primary language in Italy is Italian, which is classified as a strong FTR language. If some languages spoken had strong FTR and others weak FTR, then we classified the country as "both". For example, in Belgium, Flemish (weak FTR) and French (strong FTR) are the primary languages.

We estimate two specifications corresponding to different ways to model countries having FTR of "both". The results from these models are reported in columns 2 and 3 in Table 6. In each specification we also include an indicator for whether an individual is an Italian native or not, to allow for there to be differences in discount factors between native Italians and immigrants from other countries with strong FTR. In column 2, for countries with a FTR of "both", we assign a value of 0.5 to the strong-FTR indicator, and in column 3, we drop these observations. In line with the results in column 1, we find that native Italians have the largest discount factors. Our results are also consistent with those in Chen (2013), as we find that discount factors are strongly correlated with FTR. Immigrants from countries with weak FTR have significantly higher discount factors than those from countries with strong FTR (12 pp - 14 pp depending on the specification).

[31] See also Falk et al. (2018) who find a systematic relationship between patience and FTR.

[32] This information was collected from the CIA World Factbook.

An alternative interpretation of these results is that linguistic differences across countries could be proxying for other cultural factors that lead to differences in time preference. We do not take a stand on the particular mechanism at work. Instead, we emphasize that our results indicate that the patterns across country of origin in discounting behavior among criminals in our data are consistent with other forward-looking behaviors for the population at large. This gives additional credibility to our results and suggests that they reflect patterns in the general population as well.

Lastly, in addition to examining differences in time preference based on individual characteristics, we also examine whether there are differences across crime types. In Table 7, we report results from 10 different specifications (one for each crime category) in which we interact the discount factor with an indicator for whether the person was originally imprisoned for a crime of that type. For most crime categories, the estimated differences in discount factors are small, both statistically and economically. However, for two crime categories we find large statistically significant differences: prostitution and drug-related.

For prostitution crimes, we find a discount factor of 0.92 compared to 0.72 for all other crimes. In Italy, prostitution is legal, but organized prostitution is not. Therefore prostitution crimes in our data are those relating to the supply of prostitutes. It is interesting that those individuals participating in this type of crime have discount factors closer to what we typically assign to economic agents, as these crimes involve running a business (albeit an illegal one), as opposed to the theft of property or acts of violence.[33]

We find discount factors that are 10 percentage points lower for violations of the law regarding the use and selling of drugs. Selection into the use of drugs by a pool of more myopic individuals might explain this result. It may also be the case that drug usage directly affects time preferences and/or decision making.

5.4 Robustness Checks

In this section we discuss a number of robustness checks to our baseline specification. Estimates from these alternative specifications are provided in Tables 8 - 12.

5.4.1 Year at Entry, Age at Entry, and Time Served Effects

A potential threat to identification is that conditional on original sentence length and age at release from prison, the residual sentence is correlated with the age at entry and date at entry into prison for the pardoned sentence.[34] It may be the case that, on average, individuals who enter prison at peak offending ages have

[33]Of course this is not to say that such crimes do not involve violence, but rather that violence is not the primary activity.

[34]For example, for two individuals who are released at the same age and who were given the same original sentence, the one with the longer residual sentence entered prison at an older age and later date.

a lower propensity to re-offend, compared to individuals who enter prison at ages at which crime is less prevalent. The date of entry into prison could be related to the likelihood of recidivism if, for example, poor economic conditions lead individuals to offend, who would have otherwise not committed a crime under better conditions. In addition, the longer the residual sentence is, the less of the original sentence that was served by the offender. To the extent that this additional time served has a causal effect on the propensity to re-offend, this could bias our estimates of the discount factor.[35]

A direct way to control for these factors would be to add these variables (age at entry, date at entry, and time served) to our empirical model, specifically to our function h, to control for their potential effect. However, conditional on original sentence and age at release, these additional variables are perfectly collinear with the residual sentence.[36]

In order to address these concerns, we supplement our main dataset with additional information that allows us to control for these other factors. First, we collected data from ISTAT (the Italian National Institute of Statistics) on local GDP per capita and local unemployment rates at the time of entry into prison for each offender. Second, in order to control for potential age of entry and time served effects, we obtained data on the full incarceration history of inmates prior to the pardon. From these data we compute age of first entry into prison and total time ever served in prison, across all incarceration spells for each individual. This additional information on inmates' criminal activity prior to the incarceration spell from which they were pardoned, allows us to separately vary residual sentence, age at first incarceration, and total time ever served in prison.

For age at entry, we directly add it to our h function, as with the economic controls discussed above. For total time served, however, this approach raises another problem. Conditional on total time served, variation in residual sentence generates variation in prior time served, and individuals with more prior time served are probably also more likely to recidivate. Therefore, we need to control for the (unobserved) prior propensity to offend that is associated with this prior incarceration time. We estimate this propensity using a factor model with three measures for this prior propensity to offend, which we label ν^p.[37] We then include this in the empirical model to control for the unobserved prior propensity to offend. The details of this procedure

[35]The sign of the potential effect of time served on recidivism, however, is ambiguous. For example, additional time served could increase the likelihood of recidivism if there are criminogenic effects to incarceration (Bayer, Hjalmarsson, and Pozen (2009)) or decrease the likelihood if there is specific deterrence (Hjalmarsson (2009b)) or rehabilitation effects (Bhuller et al. (2019)).

[36]This concern was raised by Philip Cook (see Durlauf and Nagin (2011)) in reference to the results of Drago, Galbiati, and Vertova (2009) in which they find that longer residual sentences lead to a reduction in recidivism due to deterrence. He suggested that the reduction in recidivism could instead be due to the fact that individuals with shorter residual sentences, served more of their original sentence in prison, thereby increasing their propensity to re-offend. Our results in this section show that this alternative explanation is not likely to be driving the results.

[37]The three measures we use are prior time served, number of prior incarcerations, and average time between incarcerations. All of these measures are obtained from our supplemental data on incarceration histories.

are explained in Appendix C, and the factor model estimates are presented in Table C1.

In the first two columns of Table 8 we present results from specifications in which we control for the economic conditions at the time at which each offender entered prison for the incarceration associated with the pardon. In both cases, the estimated discount factors of 0.74 and and 0.75 are very similar to our baseline estimate of 0.74, which suggests that the potential selection of offenders based on economic conditions does not seem to be affecting our estimates.

In columns 3-5 of Table 8 we control for the effects of age at first incarceration and total time served in prison. Using the additional data obtained from the Italian Prison Administration, we were able to obtain incarceration spells for over 90% of our original sample. Of these, about two-thirds had prior incarceration spells, leaving us with 10,970 observations out of our original 19,616. In column 3, we first report estimates of our baseline specification (see column 2 of Table 4) on this smaller sample. The estimated discount factor of 0.68 is slightly smaller than, but in line with, our baseline estimate of 0.74. The smaller discount factor estimated here is consistent with the idea that individuals with lower discount factors are more likely to offend, and thus more likely to have a prior incarceration history. In column 4 we report estimates controlling for age at first incarceration. In column 5 we control for total time served, as well as our measure of the unobserved prior propensity to offend.[38] In both cases the estimated discount factors (0.67 and 0.70) are very similar to the baseline specification estimate in column 3 (0.68), again suggesting that the omitted variables are not biasing our results.[39]

5.4.2 Clearance Rates

Our main dataset does not contain any information on clearance rates. In order to determine whether variation in clearance rates, across either location or crime type, could affect our estimates of discounting, we perform two sets of robustness checks. In the first set we attempt to control for regional differences in clearance rates using location-specific and crime-specific dummy variables. The results are reported in columns 1-4 of Table 9. In the first column we model the clearance rates using a set of three dummy

[38] See Appendix Table C1 for the estimates from our factor model for the unobserved prior propensity to offend, and Appendix Table C2 for the coefficients on the factor and and total time served.

[39] Our finding that not controlling for these variables in our baseline results does not affect our estimates, should perhaps not be that surprising. Regarding economic conditions at entry, our dataset contains individuals with various sentence lengths and who were released at various ages. Therefore, there is no clear pattern regarding the economic conditions facing those with longer versus shorter residual sentences. Regarding the other two variables, recall that residual sentence length varies from 0 to 36 months, with an average of 15 months. Therefore, differences in age at entry into prison due to differences in the residual sentence length are small. While there may be important differences in criminality based on larger differences in age at first entry into prison, differences of a year or two are likely to be small.

A similar story exists regarding controlling for time served in prison. On average, the residual sentence represents only a quarter of the total time served for these offenders. Moreover, the variation in total time served that is induced by variation in residual sentences corresponds to the last months of total time served. It is likely that any treatment effects of incarceration accumulate most significantly during the first periods of incarceration, as opposed to the last periods.

variables, corresponding to each of the three areas of Italy (North, Center, and South). In column 2 we add crime-type dummies. In columns 3 and 4 we repeat the exercise using region dummies (there are 20 regions in Italy). The coefficients on the dummy variables are consistent with clearance rates being the highest in the North, and the estimated discount factors, which range from 0.70-0.74, are extremely similar to our baseline estimate.

We also collected separate data by province on both the total number of crimes and the number of crimes in which the perpetrator is known, using data from ISTAT (2005), to compute clearance rates. It is important to note that identifying a suspect and clearing the crime via arrest are not necessarily the same, and therefore these data serve only as proxies for the true clearance rates. Quite surprisingly, the richest regions, e.g., Lombardy, Piedmont, Veneto, Lazio, and Liguria appear to have, by far, the lowest average clearance rates. One possible explanation is that reporting rates of crimes are different across regions. The criminal justice system is widely recognized to function less efficiently in the South of Italy (see Jappelli, Pagano, and Bianco (2005)), which could lead to people being less willing to report crime[40], and thus inflating the clearance rates. We are therefore cautious not to rely too heavily on these estimates.[41] We first estimate our baseline model using the province-specific average clearance rate, and then estimate a version using the province and crime-specific clearance rate. In columns 5 and 6 we report estimates from the models using the clearance rate data. The point estimates of the discount factor increase slightly (to 0.81) in these specifications, but so does the standard error, and the resulting confidence intervals almost completely cover that of the baseline model. Overall, we conclude that accounting for heterogeneity in clearance rates has at most a small impact on our estimates of discounting.

5.4.3 Transitioning to Different Crimes

In this section we discuss estimates of the model described in Section 3.2, in which we take into account that the next crime a pardoned individual commits may differ from the one for which he was previously incarcerated, and thus carry a different sentence length. Using data on the prison population from two prisons in Milan, we constructed transition probabilities between crimes. For cases in which multiple crime types are committed, we categorized crimes based on the most serious offense (defined based on either the mean or median sentence). We constructed predicted sentences for all inmates over all possible future crime choices based on equation (3). We then integrated the likelihood over these different possible future

[40] For most property crimes, the most recent victimization survey shows that reporting rates are approximately 10 percentage points lower in the South than in the North (see Muratore et al. (2004)).

[41] Cook (1979) highlights the endogenous nature of clearance rates (they are mediated by the choices made by criminals) and advises against using these as a measure of criminal justice system effectiveness.

sentences, using the transition probabilities as weights. The estimates for this model are reported in Table 10. The results are similar, both to each other, and to our baseline estimates. The discount factor is 0.76 using the mean sentence to define the most serious crime, and 0.75 using the median. Not surprisingly the standard errors increase slightly, but the results are still fairly precisely estimated. Overall, we conclude that using the original crime type as a proxy for the future crime does not bias our estimates of time preference.

5.4.4 Sentences Less Than Two Years

Under the Collective Clemency Bill passed in 2006, only individuals who receive a future sentence of at least two years are subject to serving their residual sentence. About one-third of the individuals in our sample had original sentences of less than two years and may not face the residual sentence enhancement.

In order to investigate whether this two-year cutoff is driving any of our results, we estimated three additional specifications that deal with this issue in different ways. First, we simply set the residual sentence equal to zero for individuals with an original sentence below two years. One potential problem with this approach is that a common reason for shorter sentences is sentence reductions, and repeat offenders are less likely to obtain sentence reductions. Second, we set the new sentence equal to two years (plus the residual sentence). Third, we dropped all observations with an original sentence below two years. The results are presented in Table 11. For the first two specifications, the results are very similar to our baseline estimates, with estimated discount factors of 0.77 and 0.74. In the third specification, the point estimate drops to 0.66, but it is not nearly as precisely estimated, due to the smaller number of observations. The resulting confidence interval completely covers that of the baseline model. These results indicate that our estimates are not being driven by the two-year threshold for the law.

5.4.5 Hyperbolic Discounting

Since criminal acts generate immediate rewards but delayed costs that are spread over time, it has been argued that present bias, or a tendency to focus on the immediate gratification, might explain the engagement in criminal activities (Jolls, Sunstein, and Thaler (1998)). This kind of impatience, which is very strong for near rewards but later declines over time, is often referred as “hyperbolic discounting”.[42] While there is growing evidence that at least some individuals exhibit inconsistent time preferences (DellaVigna (2009)), for criminal behavior such evidence is still mainly speculative, with the possible exception of very young offenders (Lee and McCrary (2017)). In order to investigate the presence of time inconsistency in prefer-

[42] An overview of hyperbolic discounting can be found in Rabin (1998). In the criminology literature, a related notion is referred to as impulsivity (Wilson and Herrnstein (1985); Jolliffe and Farrington (2009)).

ences among criminals, we estimate a version of our model under hyperbolic discounting. For mathematical convenience we model hyperbolic preferences in continuous time, where such preferences are commonly expressed as

$$\delta_H(t) = \frac{1}{1+kt}, \tag{7}$$

where $\delta_H(t)$ denotes the discount factor t periods into the future. The parameter k governs the rate of discounting. In the continuous-time formulation, hyperbolic discounting differs from the exponential version in two ways. First, under the hyperbolic model, individuals discount the future to a much larger extent early on. Second, the rate of decay of discount factors is slower, leading to larger discount factors (relative to exponential discounting) in later periods. Although we have little data with short sentences, we do have data with longer sentences, which allow us to evaluate the hyperbolic model relative to the exponential model.

For convenience, researchers have often used a quasi-hyperbolic discounting specification, which is a discrete-time approximation to the continuous-time version. These functions have been adopted for analytic tractability by Laibson (1997); O'Donoghue and Rabin (1999); DellaVigna and Paserman (2005); Mastrobuoni and Weinberg (2009), among others. With quasi-hyperbolic discounting, discount factors fall quickly in the first period (or periods), and then afterward decay at a geometric rate. Under the specification due to Laibson (1997) the discount factors are given by:

$$\delta_{QH}(t) = \beta \times \delta^t,$$

where $\beta < 1$ denotes the initial drop in the discount factor. After the first period, discount factors continue to decline in the same way as under exponential discounting. Due to the nature of our data, we do not observe decisions for individuals facing extremely short sentences of only a month (or even a few months). As a result, we are not able to separately identify β from the mean disutility of prison, and the quasi-hyperbolic discounting model is observationally equivalent to the exponential discounting model for our analysis.[43] Instead, we focus on the continuous-time formulation.

In Table 12, we provide estimates from the hyperbolic form of discounting in equation (7). For comparison we also estimate a continuous time version of exponential discounting, which is given by

$$\delta_{Exp-Cont}(t) = e^{-rt},$$

[43]From a policy perspective, separately identifying the quasi-hyperbolic parameter β from the mean disutility of prison may not even matter, as long as one is not interested in the effects of very short incarceration spells, as both models would generate the same implications for utility, and therefore behavior, over such horizons.

where r is the monthly discount rate. (The estimates for the discrete and continuous time versions of exponential discounting yield results that are essentially identical to each other.)

Comparing the likelihood values in Table 12, we can see that the model with exponential discounting outperforms the model with hyperbolic discounting. Furthermore, the hyperbolic discounting parameter k is not precisely estimated. When k equals zero in equation (7), this implies no discounting, and the 95% confidence interval easily covers zero. The top end of the confidence interval implies a discount factor in the first year ($t = 12$ months) of 0.24. Together the results reflect a very large range of potential discount factors under hyperbolic discounting that are consistent with the data.

Because the parameters of the two different forms of discounting (k and r) affect discount rates in different ways, it is not straightforward to compare the implied discount factors. In order to facilitate a comparison, we have computed the implied discount factors under both estimated models and plotted them in Figure 5. One feature of hyperbolic discounting is that discount factors fall more quickly in early periods, but then flatten out, compared to exponential discounting. The implied 1-year discount factors are 0.74 and 0.44 for the exponential and hyperbolic models, respectively. At about 8 years, they intersect, with hyperbolic discount factors being higher afterward. Similar to several recent papers in the literature on estimating time preference (e.g., Warner and Pleeter (2001); Harrison, Lau, and Williams (2002); Sutter et al. (2013)), we find no evidence of hyperbolic discounting behavior, though we cannot rule out that the discount function might be hyperbolic over very short periods, giving rise to "quasi-hyperbolic" discount functions.

6 Implications for Estimates of Deterrence

Our finding of significant discounting behavior among criminals implies that the marginal deterrent effect of imprisonment is decreasing in sentence length. Estimates that fail to take this into account capture an "average" effect that is biased downward for early periods and upward for later periods. This is particularly important to consider when comparing the effectiveness of deterrence across countries and policies. For example, the elasticity of recidivism with respect to sentence length has been found to vary considerably in the literature, suggestive of large differences in deterrent power. To illustrate the importance of discounting in driving estimates of deterrence, we computed this elasticity both for the Italian pardon in our data and for the estimates in Helland and Tabarrok (2007), a study of the deterrent effect of three strikes legislation in the US. We find elasticities of -0.45 and -0.055, respectively.[44,45] The fact that the elasticity is almost an order of magnitude larger for the Italian pardon is, at face value, suggestive of large differences in the deterrent power of increases in prison sentences between Italy and the US. However, when we simulate the effect of the three strikes policy in our data using our estimated model[46], the predicted drop in recidivism implies an elasticity of only -0.10, much closer the Helland and Tabarrok (2007) estimates.

While both policies led to percentage decreases in recidivism of about 17%, the large apparent difference in deterrence is caused by the fact that the sentence enhancements under three strikes are substantially longer, as opposed to reflecting differences in preferences among offenders. The percentage increase in sentence length was 37% in the Italian pardon, and 313% under three strikes. However, the later months of the sentence are heavily discounted, leading to small incremental effects on recidivism. Using our baseline estimates of the discount factor, the percentage changes in present discounted (dis)utility from imprisonment are 18% and only 24% for the pardon and three strikes, respectively. In other words, the true force of the three strikes policy is only marginally larger than that of the Italian pardon.

For a baseline sentence s and an additional sentence Δs, we can use equation (2) to compute the percentage change in present discounted disutility of prison, $\%\Delta\text{disutility}(\delta) = \frac{\left(1-\delta^{s+\Delta s}\right)-\left(1-\delta^{s}\right)}{1-\delta^{s}}$. Compared to the elasticity with respect to sentence length, the only additional information that is needed to compute the elasticity with respect to the total disutility of prison, $\frac{\%\Delta\text{recidivism}}{\%\Delta\text{disutility}(\delta)}$, is the discount factor δ. Using our discount factor, increasing the disutility of prison by 10% reduces recidivism by 9.3% in our data and by 7.2% in Helland and Tabarrok (2007).

This analysis may also help to explain why the literature has failed to find systematic convincing evidence as to the deterrent effect of the severity of punishment (compared to certainty of punishment). The magnitude of the effect depends on the length of punishments studied, and, for long sentences, the effect is likely to be quite small after taking into account discounting.

7 Conclusion

A large number of studies have cited excessive discounting as a potential explanation for both engagement in crime as well as the (lack of) responsiveness to increased severity of punishment. Empirical evidence

[44] We compute all elasticities using the following formula:
$\frac{\%\Delta\text{recidivism}}{\%\Delta\text{sentence}} = \frac{\text{total change in recidivism rate}}{\text{baseline recidivism}} \frac{\text{baseline sentence}}{\text{months of additional sentence}}$.

[45] Drago, Galbiati, and Vertova (2009) report an elasticity of -0.74. The difference is due to the fact that, in order to facilitate comparison with other results in the literature, we compute the elasticity with respect to the baseline sentence (39 months) and the baseline recidivism (0.14), as opposed to values that include the changes in sentence and recidivism induced by the policy. If we follow their formula, we obtain an almost identical estimate to theirs of -0.75.

[46] That is, we set all future sentences to be equal to 20 years (the minimum sentence under three strikes).

on the discounting behavior of criminals, on the contrary, is quite scarce. We provide new evidence on the extent of discounting among criminals by exploiting a quasi-experiment that is driven by a mass release of Italian prison inmates that took place in 2006. Conditional on the original sentence, the collective pardon generates a distribution of exogenous sentence enhancements. The raw data show a monotonic reduction in marginal deterrence as the sentence increases (thereby reducing the additional disutility of incarceration), suggestive of the presence of discounting. In order to identify discount factors for the large number of pardoned inmates, we estimate a basic intertemporal model of criminal behavior, and our findings are very robust to a series of different specifications.

Overall, our baseline estimate of criminal discount factors of 0.74 is on the low end of the range of individual-level discount factors found in the literature for the general population. This supports the hypothesis that low future time preference is a driver of criminal behavior. However, the estimates are still far from zero, suggesting that imprisonment does have the potential to deter crime, as criminals do not fully discount future punishments.

While it has been recognized that low future time preference could be important in shaping the behavioral responses to future punishments for crimes, this is rarely taken into account both in empirical studies on deterrence and in policy debates. Given the delayed nature of imprisonment, discounting and deterrence go hand-in-hand. As we highlight above, this can have important implications for estimates of the deterrent effect of imprisonment, and can help explain the lack of consensus on the magnitude of deterrence. Correspondingly, our results imply that dynamic models of criminal behavior should employ discount factors considerably smaller than the more conventional values of 0.95-0.99 that are typically used.

Our paper provides one of the first empirical investigations and finds significant, but not complete, discounting and illustrates that accounting for discounting is critical for designing optimal deterrence policy. Our estimates suggest that while increased sentence length can have quite strong deterrent effects for low initial sentence lengths, the incremental effects are much smaller for longer sentence lengths. So while the costs of incarceration are linear in sentence length, the deterrence benefits are not. From a deterrence perspective it might be beneficial to trade off the costs of incarceration associated with long prison sentences with policies aimed at increasing the certainty of punishment either through a higher probability of apprehension or through improved efficiency of the criminal justice system.

Finally, while we have focused most of our discussion of the importance of discounting for deterrence, our results have implications for many other crime-related policies, including rehabilitation programs, education and job training programs, drug treatment programs, interventions for at-risk youth, and parole.

Any policy that involves delayed rewards or punishments needs to take into account how potential criminal offenders discount the future, in order to determine how individuals will respond to various policies. The costs and benefits need to be appropriately structured over time in order to incentivize individuals to invest effort into these programs and thus maximize their effectiveness. For example, in the context of prison educational and job training programs, the prospect of shortened sentences or improved future job market outcomes after release from prison may be ineffective due to the relatively shorter time horizons of inmates. Instead the possibility of more immediate rewards (more family visits, day release, or increased privileges) for active program participation may be more effective.

References

Åkerlund, David, Bart H.H. Golsteyn, Hans Grönqvist, and Lena Lindahl. 2016. "Time Discounting and Criminal Behavior." *Proceedings of the National Academy of Sciences* 113 (22):6160–6165.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2014. "Discounting Behavior: A Reconsideration." *European Economic Review* 71:15–33.

Barbarino, Alessandro and Giovanni Mastrobuoni. 2014. "The Incapacitation Effect of Incarceration: Evidence from Several Italian Collective Pardons." *American Economic Journal: Economic Policy* 6 (1):1–37.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen. 2009. "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections." *The Quarterly Journal of Economics* 124 (1):105–147.

Beccaria, Cesare. 1764. *On Crimes and Punishments*. Philadelphia: Philip H. Nicklin, 1819, 2nd american edition. translated from the italian by anonymous ed.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2):169–217.

Becker, Gary S. and Casey B. Mulligan. 1997. "The Endogenous Determination of Time Preference." *The Quarterly Journal of Economics* 112(3):729–758.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2019. "Incarceration, Recidivism, and Employment." *Journal of Political Economy* , forthcoming.

Blevins, Jason R., Wei Shi, Donald R. Haurin, and Stephanie Moulton. 2017. "A Dynamic Model of Reverse Mortgage Borrower Behavior." Working paper, Ohio State University.

Bukstel, Lee H. and Peter R. Kilmann. 1980. "Psychological Effects of Imprisonment on Confined Individuals." *Psychological Bulletin* 88 (2):469.

Cagetti, Marco. 2003. "Wealth Accumulation over the Life Cycle and Precautionary Savings." *Journal of Business & Economic Statistics* 21 (3):339–353.

California Assembly. 1968. "Deterrent Effects of Criminal Sanctions-Progress Report of the Assembly Committee on Criminal Procedure."

Chen, M. Keith. 2013. "The Effect of Language on Economic Behavior: Evidence from Saving Rates, Health Behaviors and Retirement Access." *American Economic Review* 103(2):690–731.

Ching, Andrew and Matthew Osborne. 2019. "Identification and Estimation of Forward-Looking Behavior: The Case of Consumer Stockpiling." *Management Science* , forthcoming.

Coller, Maribeth and Melonie B. Williams. 1999. "Eliciting Individual Discount Rates." *Experimental Economics* 2 (2):107–127.

Cook, Philip J. 1979. "The Clearance Rate as a Measure of Criminal Justice System Effectiveness." *Journal of Public Economics* 11 (1):135–142.

———. 1980. "Research in Criminal Deterrence: Laying the Groundwork for the Second Decade." *Crime and Justice* 2:211–268.

Davis, Michael L. 1988. "Time and Punishment: An Intertemporal Model of Crime." *Journal of Political Economy* 96 (2):383–90.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2):315–72.

DellaVigna, Stefano and Daniele M. Paserman. 2005. "Job Search and Impatience." *Journal of Labor Economics* 23 (3):527–587.

Dohmen, Thomas, Benjamin Enke, Armin Falk, David Huffman, and Uwe Sunde. 2015. "Patience and The Wealth of Nations." Working Paper.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova. 2009. "The Deterrent Effects of Prison: Evidence from a Natural Experiment." *Journal of Political Economy* 117 (2):257–280.

Durlauf, Steven N. and Daniel S. Nagin. 2011. "The Deterrent Effect of Imprisonment." University of Chicago Press, 43–94.

Ehrlich, Isaac. 1973. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy* 81 (3):521–565.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4):1645–1692.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.

Galor, Oded and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106 (10):3064–3103.

Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel. 2015. "Very Long-Run Discount Rates." *The Quarterly Journal of Economics* 130 (1):1–53.

Gullone, Eleonora, Tessa Jones, and Robert Cummins. 2000. "Coping Styles and Prison Experience as Predictors of Psychological Well-Being in Male Prisoners." *Psychiatry, Psychology and Law* 7 (2):170–181.

Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutström. 2010. "Individual Discount Rates and Smoking: Evidence from a Field Experiment in Denmark." *Journal of Health Economics* 29 (5):708–717.

Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. 2002. "Estimating Individual Discount Rates in Denmark: A Field Experiment." *American Economic Review* 92 (5):1606–1617.

Hausman, Jerry A. 1979. "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables." *Bell Journal of Economics* 10 (1):33–54.

Hawken, Angela and Mark Kleiman. 2009. "Managing Drug Involved Probationers with Swift and Certain Sanctions: Evaluating Hawaii's HOPE: Executive Summary." *Washington, DC: National Criminal Justice Reference Services* .

Heckman, James J. 1976. "A Life-Cycle Model of Earnings, Learning, and Consumption." *Journal of Political Economy* 84 (4):S11–44.

Heckman, James J. and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics* 30 (1-2):239–267.

Helland, Eric and Alexander Tabarrok. 2007. "Does Three Strikes Deter?: A Nonparametric Estimation." *Journal of Human Resources* 42 (2):157–188.

Hjalmarsson, Randi. 2009a. "Crime and Expected Punishment: Changes in Perceptions at the Age of Criminal Majority." *American Law and Economics Review* 11 (1):209–248.

———. 2009b. “Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?” *The Journal of Law and Economics* 52 (4):779–809.

Hochstetler, Andy, Daniel S. Murphy, and Ronald L. Simons. 2004. “Damaged Goods: Exploring Predictors of Distress in Prison Inmates.” *Crime & Delinquency* 50 (3):436–457.

Hunt, Albert R. 2015. “Reforming the Criminal Justice System Is Not Assured.” *New York Times* .

ISTAT. 2005. “Statistiche Giudiziarie Penali.” Tech. rep., Istituto Italiano di Statistica.

———. 2015. “I detenuti nelle carceri italiane.” Report, Istituto Nazionale di Statistica. URL `http://www.istat.it/it/archivio/153369`.

Italian Ministry of Internal Affairs. 2007. *Rapporto sulla criminalità in Italia. Analisi, Prevenzione, Contrasto*.

Jappelli, Tullio, Marco Pagano, and Magda Bianco. 2005. “Courts and Banks: Effects of Judicial Enforcement on Credit Markets.” *Journal of Money, Credit and Banking* 37 (2):223–244.

Jolliffe, Darrick and David P. Farrington. 2009. “A Systematic Review of the Relationship Between Childhood Impulsiveness and Later Violence.” In *Personality, Personality Disorder, and Violence*, edited by Mary McMurran and Richard C. Howard. New York: Wiley, 41–61.

Jolls, Christine, Cass R. Sunstein, and Richard Thaler. 1998. “A Behavioral Approach to Law and Economics.” *Stanford Law Review* 50:1471–1550.

Kaplow, Louis. 1990. “Optimal Deterrence, Uninformed Individuals, and Acquiring Information About Whether Acts are Subject to Sanctions.” *Journal of Law, Economics, & Organization* 6 (1):93–128.

Katz, Lawrence, Steven D. Levitt, and Ellen Shustorovich. 2003. “Prison Conditions, Capital Punishment, and Deterrence.” *American Law and Economics Review* 5 (2):318–343.

Keane, Michael P. and Kenneth I. Wolpin. 1997. “The Career Decisions of Young Men.” *Journal of Political Economy* 105 (3):473–522.

Laibson, David I. 1997. “Golden Eggs and Hyperbolic Discounting.” *The Quarterly Journal of Economics* 112 (2):443–477.

Lawrance, Emily C. 1991. “Poverty and the Rate of Time Preference: Evidence from Panel Data.” *Journal of Political Economy* 99 (1):54–77.

Lee, David S. and Justin McCrary. 2017. “The Deterrence Effect of Prison: Dynamic Theory and Evidence.” In *Regression Discontinuity Designs: Theory and Applications*. Emerald Publishing Limited, 73–146.

Levitt, Steven D. and Thomas J. Miles. 2007. “Empirical Study of Criminal Punishment.” *Handbook of Law and Economics* 1:455–495.

Lochner, Lance J. 2007. “Individual Perceptions of the Criminal Justice System.” *American Economic Review* 97 (1):444–460.

Lochner, Lance J. and Enrico Moretti. 2004. “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports.” *American Economic Review* 94 (1):155–189.

Mancino, Maria Antonella, Salvador Navarro, and David A. Rivers. 2016. “Separating State Dependence, Experience, and Heterogeneity in a Model of Youth Crime and Education.” *Economics of Education Review* 54:274–305.

Mastrobuoni, Giovanni and Paolo Pinotti. 2015. “Legal Status and the Criminal Activity of Immigrants.” *American Economic Journal: Applied Economics* 7 (2):175–206.

Mastrobuoni, Giovanni and Matthew Weinberg. 2009. “Heterogeneity in Intra-Monthly Consumption Patterns, Self-Control, and Savings at Retirement.” *American Economic Journal: Economic Policy* 1 (2):163–189.

McCrary, Justin. 2010. “Dynamic Perspectives on Crime.” In *Handbook on the Economics of Crime*, edited by Bruce L. Benson and Paul R. Zimmerman. Edward Elgar Publishing, 82–108.

Meier, Stephan and Charles Sprenger. 2010. “Present-Biased Preferences and Credit Card Borrowing.” *American Economic Journal: Applied Economics* 2 (1):193–210.

Muratore, Maria Giuseppina, Isabella Corazziari, Giovanna Tagliacozzo, Alessandro Martini, Alessandra Federici, Manuela Michelini, Agostina Loconte, Roberta Barletta, and Anna Costanza Baldry. 2004. *La Sicurezza dei Cittadini. 2002 Reati Vittime, Percezione della Sicurezza e Sistemi di Protezione. Indagine Multiscopo sulle Famiglie “Sicurezza dei cittadini?- Anno 2002*. Rome: Istituto Nazionale di Statistica.

Nagin, Daniel S. 2013. “Deterrence: A Review of the Evidence by a Criminologist for Economists.” *Annu. Rev. Econ.* 5 (1):83–105.

Nagin, Daniel S. and Greg Pogarsky. 2004. “Time and Punishment: Delayed Consequences and Criminal Behavior.” *Journal of Quantitative Criminology* 20 (4):295–317.

O’Donoghue, Ted and Matthew Rabin. 1999. “Doing It Now or Later.” *American Economic Review* 89 (1):103–124.

Polinsky, A. Mitchell and Steven M. Shavell. 1999. “On the Disutility and Discounting of Imprisonment and the Theory of Deterrence.” *Journal of Legal Studies* 28 (1):1–16.

Rabin, Matthew. 1998. “Psychology and Economics.” *Journal of Economic Literature* 36 (1):11–46.

Rust, John. 1987. “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher.” *Econometrica* 55 (5):999–1033.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rüetzler, and Stefan T. Trautmann. 2013. “Impatience and Uncertainty: Experimental Decisions Predict Adolescents’ Field Behavior.” *American Economic Review* 103 (1):510–31.

Viscusi, W. Kip and Michael J. Moore. 1989. “Rates of Time Preference and Valuations of the Duration of Life.” *Journal of Public Economics* 38 (3):297–317.

Warner, John T. and Saul Pleeter. 2001. “The Personal Discount Rate: Evidence from Military Downsizing Programs.” *American Economic Review* 91 (1):33–53.

Wilson, James Q. and Richard J. Herrnstein. 1985. *Crime and Human Nature: The Definitive Study of the Causes of Crime*. New York: Simon and Schuster.

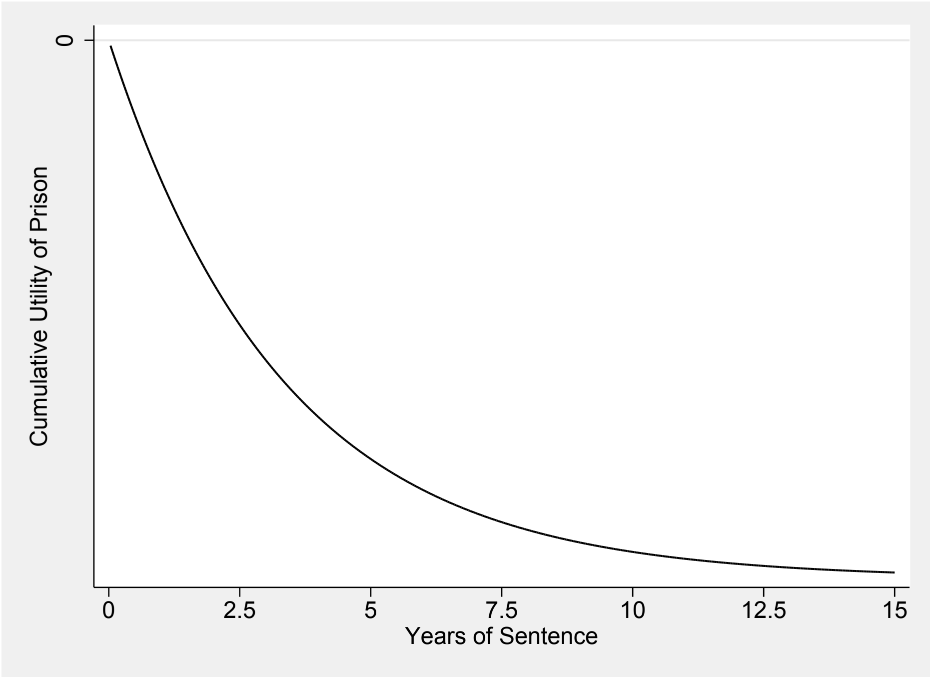

Figure 1: Cumulative Utility of Prison as a Function of Sentence Length

Notes: The figure shows the cumulative utility of prison time as a function of sentence length assuming an annual discount factor of 0.74. Note that all values of cumulative utility are negative.

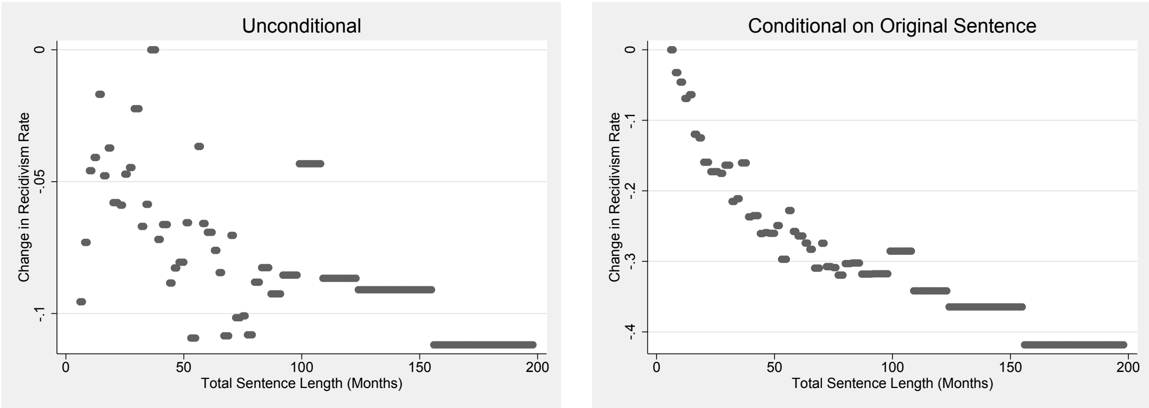

Figure 2: Change in Average Recidivism by Total Sentence Length

Notes: The left panel plots the average recidivism rate for each of 40 quantiles of the total expected sentence length. The right panel adds controls for original sentence length. In order to control for original sentence length we regress the recidivism indicator on a cubic polynomial in the log of original sentence and dummies for each of the quantiles of total sentence length. The right panel plots the coefficients on each of the dummies. For both panels the differences in average recidivism are normalized relative to the maximum recidivism rate across the quantiles.

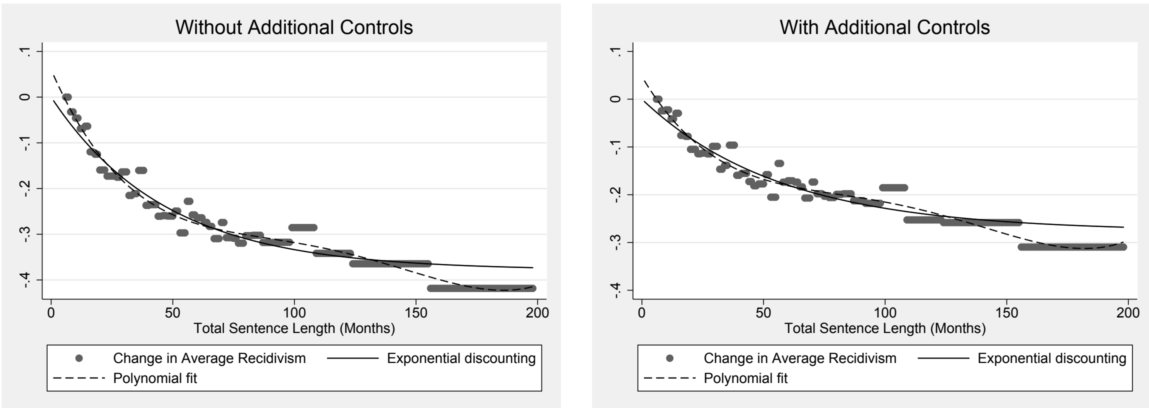

Figure 3: Change in Average Recidivism by Total Sentence Length with Fit

Notes: The left panel plots the average recidivism rate for each of 40 quantiles of the total expected sentence length, including only controls for original sentence length. The right panel adds additional controls, including indicators for marital status, gender, definitive sentence, nationality, permanent employment status, crime type, as well as measures of education and age. In order to control for original sentence length we regress the recidivism indicator on a cubic polynomial in the log of original sentence and dummies for each of the quantiles of total sentence length. In both panels we have added a fourth-degree polynomial and exponential fit to the plots.

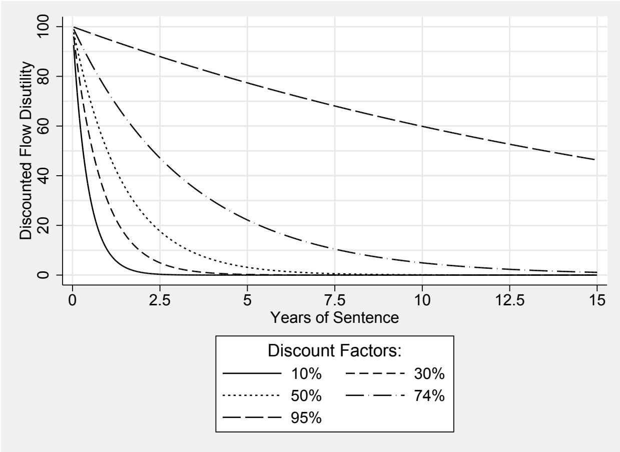

Figure 4: Flow Disutility of Prison as a Function of Sentence Length–Different Discount Factors

Notes: The figure shows the discounted flow of disutility of prison time (normalized to be 100 at the beginning of incarceration) for different annual discount factors.

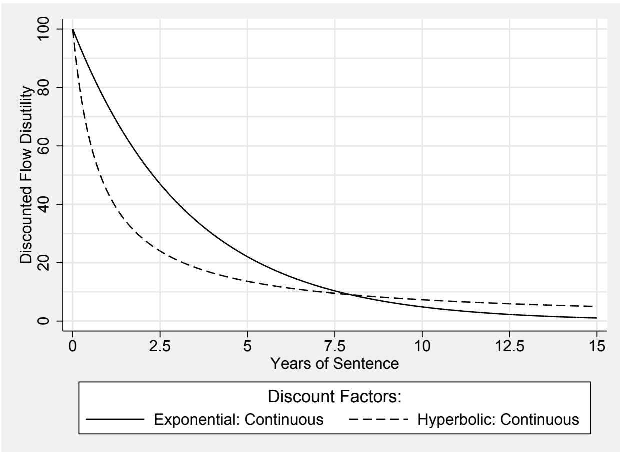

Figure 5: Flow Disutility of Prison as a Function of Sentence Length–Different Forms of Discounting

Notes: The figure shows the discounted flow of disutility of prison time (normalized to be 100 at the beginning of incarceration) for exponential and hyperbolic discounting. These curves are based on the estimates in Table 12.

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Individual Characteristics:				
Recidivism (re-incarceration)	0.22	0.41	0	1
Original sentence	39.90	32.67	1	360
Residual sentence	15.01	10.52	0	36
Definitive sentence	0.74	0.44	0	1
Age	37.72	9.79	19	70
Female	0.05	0.21	0	1
Married	0.26	0.44	0	1
Permanently employed	0.15	0.36	0	1
Primary education or less	0.27	0.44	0	1
No education	0.06	0.24	0	1
Secondary school or above	0.03	0.18	0	1
High school or above	0.02	0.13	0	1
Italian	0.62	0.49	0	1
Crime Type:				
Property	0.58	0.49	0	1
Violent	0.26	0.44	0	1
Illegal detention of weapons	0.09	0.29	0	1
Drug-related	0.39	0.49	0	1
Prostitution	0.02	0.13	0	1
Illegal migration	0.07	0.26	0	1
Organized crime (mafia)	0.02	0.13	0	1
Against the economy and state	0.26	0.44	0	1
Fraud	0.10	0.29	0	1
Other crimes	0.19	0.39	0	1
Previous Incarcerations:				
Number of incarcerations	3.67	3.05	1	30
Total prior time served	30.47	34.96	1	288
Average time free	35.15	30.01	2.16	217.95

Notes: This table presents summary statistics for our data on pardoned inmates. All durations are expressed in months. The first two panels correspond to our main sample of 19,616 observations. The third panel (previous incarcerations) corresponds to our supplemental sample of individuals for which we have data on at least one incarceration spell prior to the one associated with the pardon. This sample consists of 10,970 observations.

Table 2: Clearance Rates

Crime Type	Mean	Std. Dev.	Min.	Max.
Property	0.11	0.08	0.03	0.39
Violent	0.58	0.17	0.14	0.94
Drug-related	0.73	0.17	0.19	1.00
Organized crime (mafia)	0.88	0.16	0.00	1.00
Against the economy and state	0.50	0.11	0.21	0.83
Fraud	0.33	0.13	0.08	0.82
Other crimes	0.77	0.31	0.00	1.00
All	0.23	0.11	0.08	0.63

Notes: The province-level clearance rates are based on ISTAT's criminal statistics (ISTAT, 2005).

Table 3: Transition Probabilities

Panel A: Most Serious Crime Based on Median Sentence

	Property	Violent	Drug-related	Prostitution	Organized crime	Against economy/state	Fraud	Other crimes	Sum
Property	0.47	0.14	0.09	0.00	0.01	0.17	0.03	0.09	1
Violent	0.24	0.33	0.12	0.00	0.02	0.15	0.03	0.11	1
Drug-related	0.13	0.10	0.51	0.00	0.01	0.08	0.03	0.14	1
Prostitution	0.25	0.00	0.17	0.25	0.00	0.17	0.00	0.17	1
Organized crime (mafia)	0.10	0.14	0.10	0.00	0.37	0.13	0.04	0.12	1
Against the economy and state	0.28	0.15	0.10	0.00	0.02	0.29	0.05	0.10	1
Fraud	0.21	0.13	0.15	0.00	0.04	0.18	0.22	0.08	1
Other crimes	0.22	0.15	0.25	0.00	0.03	0.15	0.04	0.17	1

Panel B: Most Serious Crime Based on Mean Sentence

	Property	Violent	Drug-related	Prostitution	Organized crime	Against economy/state	Fraud	Other crimes	Sum
Property	0.47	0.13	0.10	0.00	0.01	0.17	0.02	0.09	1
Violent	0.24	0.32	0.13	0.01	0.02	0.15	0.02	0.11	1
Drug-related	0.13	0.10	0.52	0.00	0.01	0.08	0.02	0.14	1
Prostitution	0.22	0.09	0.09	0.13	0.00	0.17	0.04	0.26	1
Organized crime (mafia)	0.10	0.13	0.11	0.01	0.37	0.13	0.03	0.12	1
Against the economy and state	0.28	0.15	0.11	0.00	0.02	0.29	0.04	0.10	1
Fraud	0.23	0.14	0.11	0.00	0.03	0.20	0.21	0.07	1
Other crimes	0.22	0.14	0.27	0.00	0.03	0.15	0.03	0.17	1

Notes: The transition probabilities are based on the entire criminal history of a sample of inmates who served at least some time in one of two different Milan prisons between 2001 and 2012. Since criminals may be imprisoned for an offense related to multiple crime types, the transitions are based on the crimes with either the highest median sentence (Panel A) or the highest mean sentence (Panel B).

Table 4: Baseline Results

Specification	(1) No Covariates	(2) Baseline Covariates	(3) Alt. Low Education	(4) Alt. High Education	(5) Crime Dummies
Monthly Discount Factor (δ)	0.97142	0.97488	0.97546	0.97469	0.97104
	(0.00537)	*(0.00519)*	*(0.00511)*	*(0.00526)*	*(0.00549)*
Annual Discount Factor	0.706	0.737	0.742	0.735	0.703
Controls for Original Sentence:					
Log-sentence	0.366	0.252	0.174	0.279	0.372
	(0.409)	*(0.398)*	*(0.393)*	*(0.399)*	*(0.425)*
Log-sentence sq.	0.199	0.174	0.203	0.163	0.133
	(0.145)	*(0.146)*	*(0.146)*	*(0.147)*	*(0.152)*
Log-sentence cu.	-0.032	-0.026	-0.028	-0.025	-0.025
	(0.016)	*(0.016)*	*(0.016)*	*(0.016)*	*(0.017)*
Utility of Crime:					
Constant	-1.841	-1.169	-1.144	-1.190	-1.153
	(0.379)	*(0.439)*	*(0.435)*	*(0.440)*	*(0.459)*
Age		-0.010	-0.010	-0.009	-0.010
		(0.006)	*(0.006)*	*(0.006)*	*(0.006)*
Permanently employed		-0.331	-0.336	-0.333	-0.337
		(0.195)	*(0.193)*	*(0.196)*	*(0.209)*
Primary education or less		-0.024		-0.028	-0.013
		(0.123)		*(0.124)*	*(0.131)*
No education			0.333		
			(0.216)		
Secondary education and above		-0.511	-0.467		-0.562
		(0.361)	*(0.356)*		*(0.387)*
High school and above				-0.568	
				(0.441)	
Italian		0.026	0.055	0.025	0.065
		(0.114)	*(0.115)*	*(0.115)*	*(0.136)*
Female		-0.405	-0.429	-0.412	-0.443
		(0.288)	*(0.286)*	*(0.289)*	*(0.306)*
Married		-0.597	-0.609	-0.600	-0.613
		(0.137)	*(0.136)*	*(0.138)*	*(0.147)*
Crime dummies					√
Disutility of Prison:					
Constant	-0.068	-0.035	-0.034	-0.035	-0.051
	(0.018)	*(0.014)*	*(0.014)*	*(0.014)*	*(0.019)*
Age		-0.001	-0.001	-0.001	-0.001
		(0.000)	*(0.000)*	*(0.000)*	*(0.000)*
Permanently employed		0.001	0.001	0.000	0.002
		(0.007)	*(0.006)*	*(0.007)*	*(0.008)*
Primary education		0.007		0.008	0.007
		(0.005)		*(0.005)*	*(0.005)*
No education			-0.006		
			(0.008)		
Secondary education and above		0.007	0.004		0.011
		(0.013)	*(0.012)*		*(0.015)*
High school and above				0.010	
				(0.018)	
Definitive sentence		-0.014	-0.014	-0.014	-0.012
		(0.002)	*(0.002)*	*(0.002)*	*(0.002)*
Italian		0.020	0.019	0.020	0.014
		(0.005)	*(0.005)*	*(0.005)*	*(0.006)*
Female		-0.020	-0.019	-0.020	-0.018
		(0.012)	*(0.011)*	*(0.012)*	*(0.013)*
Married		0.012	0.012	0.012	0.014
		(0.005)	*(0.005)*	*(0.005)*	*(0.006)*
Crime dummies					√
Obs.	19616	19616	19616	19616	19616
log-likelihood	-10197.64	-9846.72	-9854.02	-9848.48	-9746.41

Notes: This table contains estimates from our baseline model in equation (6). In column 1 we control only for original sentence using a polynomial in log of original sentence. In column 2 we add our baseline controls. Columns 3 and 4 use alternative measures of education, and column 5 allows the utility of committing a crime, as well as the disutility of prison, to depend on a full set of crime-type dummies. Standard errors are reported in parentheses below the point estimates.

Table 5: Heterogeneity in Discount Factors

Specification:	(1) Married	(2) Female	(3) Permanently employed	(4) Log(age-18)	(5) Age, age sq.	(6) Primary education or less	(7) No education	(8) Secondary education and above	(9) High school and above
Monthly Discount Factor (δ)									
Constant	0.97280	0.97390	0.97506	0.99526	1.00842	0.97508	0.97567	0.97456	0.97440
	(0.00547)	*(0.00523)*	*(0.00547)*	*(0.00837)*	*(0.01459)*	*(0.00527)*	*(0.00519)*	*(0.00535)*	*(0.00514)*
Interaction coefficient	0.00674	0.00979	-0.00051	-0.00780	-0.00161	-0.00086	-0.00177	0.01529	0.02520
	(0.00417)	*(0.00753)*	*(0.00506)*	*(0.00271)*	*(0.00063)*	*(0.00389)*	*(0.00639)*	*(0.00694)*	*(0.00892)*
Age squared					0.000016				
					(0.000006)				
Annual Discount Factor									
Baseline	0.718	0.728	0.739			0.739	0.744	0.734	0.733
With interaction	0.780	0.821	0.734			0.731	0.728	0.885	0.995
Difference	0.062	0.093	-0.005			-0.008	-0.016	0.151	0.263
At age 26				0.776	0.760				
At age 51				0.677	0.676				
Obs.	19616	19616	19616	19616	19616	19616	19616	19616	19616
log-likelihood	-9845.52	-9846.04	-9846.72	-9842.45	-9842.26	-9846.70	-9853.98	-9845.14	-9846.62

Notes: In these specifications we allow the discount factor to differ by the variables shown in the first row. Standard errors are reported in parentheses below the point estimates.

Table 6: Heterogeneity in Discount Factors

Specification:	(1) Nationality	(2) Nationality-Language 1	(3) Nationality-Language 2
Monthly Discount Factor (δ)			
Constant	0.96547	0.97551	0.97956
	(0.00872)	*(0.00977)*	*(0.00851)*
Italian	0.01442	0.01886	0.01418
	(0.00613)	*(0.00915)*	*(0.00602)*
Strong FTR		-0.01674	-0.01408
		(0.00751)	*(0.00601)*
Annual Discount Factor			
Italian	0.784	0.762	0.780
Immigrant (All)	0.656		
Immigrant (Strong FTR)		0.603	0.656
Immigrant (Weak FTR)		0.743	0.780
Obs.	19616	19571	19220
log-likelihood	-9842.21	-9821.16	-9700.85

Notes: In these specifications we allow the discount factor to differ by nationality and language. FTR refers to future time reference. There are 45 observations for which either the country of origin was unknown or we were unable to classify the FTR for the country. Standard errors are reported in parentheses below the point estimates.

Table 7: Heterogeneity in Discount Factors by Crime Type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Interaction variable:	Violent	Property	Arms	Drug-related	Prostitution	Migrant Law	Organized crime (mafia)	Against the economy and state	Fraud	Other crimes
Monthly Discount Factor (δ)										
Constant	0.97191	0.97715	0.97637	0.98190	0.97284	0.97041	0.97571	0.97529	0.97382	0.97243
	(0.00513)	*(0.00525)*	*(0.00497)*	*(0.00511)*	*(0.00530)*	*(0.00566)*	*(0.00543)*	*(0.00510)*	*(0.00507)*	*(0.00533)*
Interaction coefficient	0.00364	-0.00407	-0.00463	-0.01160	0.02052	-0.04047	-0.00825	-0.00841	-0.01133	0.00843
	(0.00385)	*(0.00595)*	*(0.00793)*	*(0.00387)*	*(0.00649)*	*(0.03629)*	*(0.01619)*	*(0.00564)*	*(0.00971)*	*(0.00409)*
Annual Discount Factor										
Overall	0.710	0.758	0.751	0.803	0.719	0.697	0.744	0.741	0.727	0.715
Difference	0.033	-0.037	-0.042	-0.107	0.205	-0.279	-0.072	-0.073	-0.095	0.078
Obs.	19616	19616	19616	19616	19616	19616	19616	19616	19616	19616
log-likelihood	-9841.44	-9805.33	-9842.98	-9837.12	-9837.18	-9841.82	-9845.78	-9800.75	-9843.39	-9836.25

Notes: In these specifications we allow the discount factor to differ by the types of crimes committed by the inmates before the pardon. Standard errors are reported in parentheses below the point estimates.

Table 8: Robustness–Year at Entry, Age at Entry, and Time Served

Specification	(1)	(2)	(3)	(4)	(5)
	Local GDP	Local Unemployment Rate	Baseline	Age at First Incarceration	Time Served
Monthly Discount Factor (δ)	0.97523	0.97639	0.96894	0.96781	0.9711859
	(0.00554)	*(0.00537)*	*(0.00645)*	*(0.00701)*	*(0.01095)*
Annual Discount Factor	0.740	0.751	0.685	0.675	0.704
Controlling for Local GDP	√				
Controlling for Local Unemployment Rate		√			
Controlling for Age at First Incarceration				√	
Controlling for Time Served					√
Obs.	19339	19481	10970	10970	10970
log-likelihood	-9712.07	-9786.73	-6145.64	-6105.19	-5885.74

Notes: In column 1 we control for economic conditions at the time of entry into prison using a polynomial in the log of local annual GDP per capita. In column 2 we control for economic conditions at the time of entry into prison using a polynomial in the log of the local quarterly unemployment rate. Local GDP and the local unemployment rate are both measured at the province level and are taken from ISTAT, the Italian National Insitute of Statistics. We lose 277 observations in column 1 and 135 in column 2 as the local GDP and unemployment data only go back to 1995 and 1993, respectively. In column 3 we report estimates of our baseline specification (see column 2 of Table 4) on the sample of individuals for which we observe a previous incarceration history prior to the sentence for which they were pardoned. In column 4 we control for the age at which each individual was first incarcerated, using a log polynomial in age. In column 5, we control for total cumulative time served, conditioning also on our measure of the prior propensity to offend. We use log polynomials in both total cumulative time served and prior propensity to offend. Standard errors are reported in parentheses below the point estimates. The standard errors in column 5 control for the fact that our measure of the prior propensity to offend is itself an estimate.

Table 9: Robustness–Various Controls for Clearance Rates

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Monthly Discount Factor (δ)	0.97291	0.97524	0.97111	0.97392	0.98263	0.98269
	(0.00515)	*(0.00551)*	*(0.00504)*	*(0.00554)*	*(0.00798)*	*(0.00729)*
Annual Discount Factor	0.719	0.740	0.703	0.728	0.810	0.811
Area dummies	$\surd$	$\surd$				
Region dummies			$\surd$	$\surd$		
Crime-type dummies		$\surd$		$\surd$		
Clearance rates by province					$\surd$	
Clearance rates by province and crime type						$\surd$
Obs.	19616	19616	19616	19616	19616	19616
log-likelihood	-9838.81	-9752.81	-9821.16	-9734.41	-9881.57	-9882.37

Notes: In column 1 we proxy for clearance rates using dummy variables for each of the three areas of Italy (North, Center, and South). In column 2 we include area dummies as well as dummies for the crime type of the original offense. In column 3 we use dummies for each region (20 regions), instead of areas. Column 4 adds crime-type dummies. In columns 5 and 6 we use province-level clearance rates (taken from ISTAT, 2005), overall and crime-specific, respectively. Standard errors are reported in parentheses below the point estimates.

Table 10: Robustness–Incorporating Transition Probabilities Across Crime Types

	(1)	(2)
Specification	Mean	Median
Monthly Discount Factor (δ)	0.97750	0.97613
	(0.00616)	*(0.00641)*
Annual Discount Factor	0.761	0.748
Obs.	19616	19616
log-likelihood	-9854.88	-9854.98

Notes: These results are based on the model in Section 3.2 in which we intergrate against the distribution of future crime choices, conditional on the original crime type. We classify offenses into the most serious crime type, based on both the mean (column 1) and median (column 2) sentence for each type. Standard errors are reported in parentheses below the point estimates.

Table 11: Robustness–The Expected New Sentence

Specification	(1)	(2)	(3)
Monthly Discount Factor (δ)	0.97860	0.97477	0.96577
	(0.00604)	*(0.00771)*	*(0.01321)*
Annual Discount Factor	0.771	0.736	0.658
Obs.	19616	19616	13241
log-likelihood	-9854.23	-9852.64	-6449.57

Notes: Each column presents a different approach to dealing with original sentences below two years. Each is based on our baseline specification. In column 1, we set the residual sentence equal to zero for observations with original sentences below two years. In column 2 we set the expected new sentence equal to two years (plus the residual sentence). In column 3, we drop these observations. Standard errors are reported in parentheses below the point estimates.

Table 12: Robustness–Alternative Forms of Discounting

	(1)	(2)
Specification	Exponential–Continuous	Hyperbolic–Continuous
r	0.02543	
	(0.00532)	
k		0.10150
		(0.07441)
Obs.	19616	19616
log-likelihood	-9846.72	-9848.08

Notes: The parameters r and k govern the rate of discounting in the exponential and hyperbolic models, respectively. Standard errors are reported in parentheses below the point estimates.

Appendix A: Dynamic Utility Model

In this appendix we show how our main estimating equation can be obtained from a more general dynamic utility model via restrictions on the option value of future crimes. The value of being free and facing sentence s if caught committing another crime is given by

$$V^f(s) = \max\left\{u(nc) + \delta E\left[V^f(s)\right], u(c) + P\left(d + \delta E\left[V^p(s-1)\right]\right) + \delta(1-P)E\left[V^f(s)\right] + \varepsilon\right\},$$

where $u(nc)$ is the utility from not committing a crime, $u(c)$ is the utility from committing a crime, P is the probability of being apprehended, $V^p(s)$ is the utility of being in prison with a sentence of s years, d is the per-period disutility of prison, and ε is a random utility term associated with committing a crime. Without loss of generality, we normalize $u(nc)$ to be equal to zero.

If an individual commits a crime and is apprehended, they are placed in prison for s total periods periods (the first of which is served immediately), and then released. Therefore the value of being in prison with a remaining sentence of $s-1$ periods can be written as:

$$V^p(s-1) = d\left(\frac{1-\delta^{s-1}}{1-\delta}\right) + \delta^{s-1}E\left[V^f(s)\right].$$

One way to generate our main estimating equation is to assume that the utility shocks ε vary across individuals, but are constant over time. Under this assumption, there will be a cutoff value of ε (which depends on observables) that determines whether or not an individual commits a crime when not incarcerated. Individuals with values of ε below the cutoff will not commit crime, which given our normalization of $u(nc) = 0$ implies that the value of their option to commit future crime is zero $\left(E\left[V^f(s)\right] = 0\right)$, and we obtain the model in equation (2). For individuals with an ε above the cutoff, we can write their value functions as follows. Since ε is fixed over time, $V^f(s) = E\left[V^f(s)\right]$, and since ε is above the cutoff, these individuals will choose to commit crime, which implies the following:

$$V^f(s) = u(c) + P\left(d + \delta V^p(s-1)\right) + \delta(1-P)V^f(s) + \varepsilon.$$

Plugging in for the value of being incarcerated $V^p(s)$, we have

$$V^f(s) = u(c) + P\left(d + \delta d\left(\frac{1-\delta^{s-1}}{1-\delta}\right) + \delta^s V^f(s)\right) + \delta(1-P)V^f(s) + \varepsilon$$

which we can solve for:

$$V^f(s) = \frac{u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon,}{1 - P\delta^s - \delta(1-P)}.$$

Since these individuals choose to commit crime it must be the case that they prefer to do it earlier rather than later (due to discounting), meaning that

$$0 + \delta\left[\frac{u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon}{1 - P\delta^s - \delta(1-P)}\right] <$$
$$u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + (P\delta^s + \delta(1-P))\left[\frac{u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon,}{1 - P\delta^s - \delta(1-P)}\right] + \varepsilon.$$

Subtracting $\delta\left[\frac{u(c)+Pd\left(\frac{1-\delta^s}{1-\delta}\right)+\varepsilon,}{1-P\delta^s-\delta(1-P)}\right]$ from both sides we have

$$0 < u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon + (P\delta^s + \delta(1-P) - \delta)\left[\frac{u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon,}{1 - P\delta^s - \delta(1-P)}\right],$$

which can be rewritten as

$$0 < \left(\frac{1-\delta}{1 - P\delta^s - \delta(1-P)}\right)\left[u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon\right].$$

For $\delta < 1$, $\left(\frac{1-\delta}{1-P\delta^s-\delta(1-P)}\right)$ is always positive, and this inequality holds if and only if $u(c) + Pd\left(\frac{1-\delta^s}{1-\delta}\right) + \varepsilon > 0$. Therefore, we have shown that for individuals with random utility draws ε both above and below the cutoff, their decisions are consistent with the model we employ in Section 3.1. If we plug in for $u(c)$ and d, and add back in subscripts, this is equivalent to $\alpha_0 + \alpha_1 X_{ij}^u + \left[\beta_0 + \beta_1 X_{ij}^d\right] P_{jl}^c \left[\frac{1-\delta\left(X_{ij}^\delta\right)^{ts_{ij}}}{1-\delta\left(X_{ij}^\delta\right)}\right] + \varepsilon_{ij} > 0$, which is the model in equation (2).

An alternative way to motivate our model is to set $E\left[V^f(s)\right] = 0$, which essentially sets the value of having the option to commit future crime to be zero. For example, suppose that individuals make a once-and-for-all decision of whether to commit a crime, and then receive zero payoff afterward. In this case the model simplifies to:

$$V^f(s) = \max\left\{0, u(c) + Pd\frac{1-\delta^s}{1-\delta} + \varepsilon\right\}.$$

We then have that an individual commits a crime if and only if $u(c) + Pd\frac{1-\delta^s}{1-\delta} + \varepsilon > 0$, which again is consistent with our model in equation (2).

Appendix B: Conditional Independence Assumption

Recall that c_{ij} is an indicator for whether individual i commits a crime of type j. For notational simplicity, we drop the subscripts and let c be an indicator for committing a crime. The probability that a crime is committed conditional on the original sentence os, the residual sentence rs, the other observables X, and a vector of parameters to be estimated θ, can be written as:

$$\begin{aligned}
\Pr(c=1 \mid os, rs, X; \theta) &= \Pr(g(os, rs, X; \theta) + \varepsilon > 0) \\
&= \Pr(\varepsilon > -g(os, rs, X; \theta)) \\
&= 1 - F_{\varepsilon|os,rs}(-g(os, rs, X; \theta)) \\
& \quad 1 - F_{\varepsilon|os}(-g(os, rs, X; \theta)),
\end{aligned}$$

where g is the utility model as a function of observables and F is the distribution function for ε. The last equality follows from the key identifying assumption that conditional on the original sentence os, the error ε is independent of the residual sentence rs.

Recognizing the dependence between ε and os, we can write

$$\varepsilon = h(os; \gamma) + u,$$

where h is an unknown (non-parametric) function. Assuming u is independent of os, and using a simple change of variables from ε to u, we can then write the probability of crime as

$$\begin{aligned}
\Pr(c=1 \mid os, rs, X; \theta) &= 1 - F_{\varepsilon|os}(-g(os, rs, X; \theta)) \\
&= 1 - F_{u|os}(-g(os, rs, X; \theta) - h(os; \gamma)) \\
&= 1 - F_{u}(-g(os, rs, X; \theta) - h(os; \gamma)),
\end{aligned}$$

where the second equality is due to the change of variables, and the third equality follows from the independence of u and os. This implies that in addition to our model of the crime choice g, an additional term needs to be added to the model: $h(os; \gamma)$. This function controls for the dependence between the error ε and the original sentence os. This additional term is necessary because we do not know the exact form of the dependence h between ε and os. It may be the case that os enters h differently from how it enters g. The model can then be estimated using maximum likelihood, with the caveat that any part of g that is separable in os is not separately identified from h. Since our only interest is in identifying the part of g related to the residual sentence, this is not a problem for identification of the discount factor.

Appendix C: Controlling for Total Time Served

Conditional on the original sentence length (os_i), the time served from the pardoned sentence does not vary separately from the residual sentence. In order to solve this, we bring in additional data on past incarceration histories, which allows us to construct a measure of the total (cumulative) time served by each individual, summed across all incarceration spells. We denote the total (cumulative) time served by individual i as cs_i, which can be decomposed as the time served prior to the pardoned sentence, denoted ps_i, and time served from the pardoned sentence $(os_i - rs_i)$. We can then vary total time served holding the residual sentence fixed, through variation in prior time served.

Ideally we would then just control for cs_i in our empirical model. However, conditional on original sentence and cumulative time served, residual sentence is only going to vary if prior time served varies. Prior time served is likely correlated with the unobserved propensity to offend prior to the incarceration associated with the pardon, and since this is likely correlated with the propensity to offend following the pardon release, we need to control for this. In order to do so, we estimate a factor model for this unobserved propensity to offend, and include an estimate of the factor in our empirical model. Below we first describe the factor model we use, and then we show how we use the resulting estimates in our estimating equation for the discount factor.

Factor Model: The factor is the unobserved propensity to offend, which we denote ν_i^p. Our measurements of the factor are (log) prior time served, number of prior incarcerations, and (log) average time between incarcerations and are denoted $x_i = (x_{1i}, x_{2i_i}, x_{3i})$. This generates the following system of equations relating the factor to each of the measurements:

$$\begin{aligned} x_{1i} &= \lambda_{10} + \lambda_{11}\nu_i^p + \varepsilon_{1i} \\ x_{2i} &= \lambda_{20} + \lambda_{21}\nu_i^p + \varepsilon_{2i} \\ x_{3i} &= \lambda_{30} + \lambda_{31}\nu_i^p + \varepsilon_{3i}, \end{aligned} \tag{8}$$

While this model is identified nonparametrically, we impose distributional assumptions in order use the estimates of this model to predict the distribution of ν_i^p for each individual. Specifically we assume that

$\left(\nu_i^p, \varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}\right)$ are each distributed normally. Without loss of generality we normalize the mean of the factor ν_i^p to one. We estimate the model using maximum likelihood. Given estimates of the $\lambda's$ and the variances of the $\varepsilon's$, we can then calculate the estimated distribution of the factor for each individual using Bayes' Rule. Let $\hat{\lambda}_0$ and $\hat{\lambda}_1$ denote the 3×1 vectors of estimated means and loadings from equation (8) above and let $\hat{\psi}$ denote a 3×3 diagonal matrix, where the diagonal elements are the estimated variances of the $\varepsilon's$: $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2$. The estimates from equation (8) are provided in Table C1.

For each individual i, the distribution of the factor, conditional on the $x's$, is normally distributed:

$$\nu_i^p \mid (x_i) \sim N\left(\hat{\lambda}_1' \left(\hat{\lambda}_1 \hat{\lambda}_1' + \hat{\psi}\right)^{-1} \left(x_i - \hat{\lambda}_0\right), 1 - \hat{\lambda}_1' \left(\hat{\lambda}_1 \hat{\lambda}_1' + \hat{\psi}\right)^{-1} \hat{\lambda}_1\right).$$

We can then add both total time served and the factor to our main estimating equation (equation (6)):

$$V\left(c_{ij} = 1\right) = \alpha_0 + \alpha_1 X_{ij}^u + \left[\beta_0 + \beta_1 X_{ij}^d\right] P_{jl}^c \left[\frac{1 - \delta\left(X_{ij}^\delta\right)^{ts_{ij}}}{1 - \delta\left(X_{ij}^\delta\right)}\right] + f\left(cs_i; \gamma^f\right) + g\left(\nu_i^p; \gamma^g\right) + h\left(os_{ij}; \gamma\right) + u_{ij},$$

where γ^f and γ^g denote the parameters of the functions f and g. We can then estimate the equation above by maximum likelihood, integrating over the distribution of the prior propensity to offend, ν_i^p. (We note that we obtain similar results evaluating the likelihood at the mean of the distribution of ν_i^p, as opposed to integrating over its distribution.)

Table C1: Factor Model Estimates

	Constant	Loading	Variance of the Error	Fraction of Variance Explained by Factor
Measurement:				
(Log) Prior Time Served	0.00	0.69	0.53	0.47
	(0.01)	*(0.01)*	*(1.02)*	–
Number of Prior Incarcerations	0.00	0.75	0.43	0.57
	(0.01)	*(0.01)*	*(1.03)*	–
(Log) Average Time Between Incarcerations	0.00	-0.59	0.66	0.34
	(0.01)	*(0.01)*	*(1.02)*	–

Notes: Without loss of generality, each of the three measurements (log prior time served, number of prior incarcerations, and log averge time between incarcerations) are normalized to have mean zero and variance one, and the mean of the factor is normalized to have mean one. Standard errors are reported in parentheses below the point estimates.

Table C2: Time Served Estimates

Monthly Discount Factor (δ)	0.97119
	(0.01095)
Annual Discount Factor	0.704
Factor (ν_i^p)	0.368
	(0.054)
Factor-squared	-0.123
	(0.029)
Factor-cubed	0.011
	(0.009)
(Log) Total Cumulative Time Served	0.241
	(0.234)
(Log) Total Cumulative Time Served–squared	-0.104
	(0.082)
(Log) Total Cumulative Time Served–cubed	0.020
	(0.009)
Average Marginal Effect of Time Served	0.537
Obs.	10970
log-likelihood	-5885.74

Notes: In this table we present the coefficients on the estimated factor and total cumulative time served for the estimates in column 5 of Table 8. Standard errors are reported in parentheses below the point estimates.