
How Much Crime Reduction Does the Marginal Prisoner Buy?
Author(s): Rucker Johnson and Steven Raphael
Source: *Journal of Law and Economics*, Vol. 55, No. 2 (May 2012), pp. 275-310
Published by: The University of Chicago Press for The Booth School of Business of the University of Chicago and The University of Chicago Law School
Stable URL: http://www.jstor.org/stable/10.1086/664073
Accessed: 25/06/2014 21:20

How Much Crime Reduction Does the Marginal Prisoner Buy?

Rucker Johnson *University of California, Berkeley*
Steven Raphael *University of California, Berkeley*

Abstract

We estimate the effect of changes in incarceration rates on changes in crime rates using state-level panel data. We develop an instrument for future changes in incarceration rates based on the theoretically predicted dynamic adjustment path of the aggregate incarceration rate in response to a shock to prison entrance or exit transition probabilities. Given that incarceration rates adjust to permanent changes in behavior with a dynamic lag, one can identify variation in incarceration rates that is not contaminated by contemporary changes in criminal behavior. For the period 1978–2004, we find crime-prison elasticities that are considerably larger than those implied by ordinary least squares estimates. We also present results for two subperiods: 1978–90 and 1991–2004. Our instrumental variables estimates for the earlier period suggest relatively large crime-prison effects. For the later time period, however, the effects of changes in incarceration rates on crime rates are much smaller.

1. Introduction

Between 1980 and 2008, the number of inmates in U.S. state and federal prisons increased from approximately 320,000 to more than 1.6 million. This corresponds to a change in the incarceration rate from 139 to 505 prisoners per 100,000 residents. It is not surprising that expenditures on corrections increased in tandem as states built new prisons, expanded corrections employment, and incurred the additional costs of housing and supervising greater numbers of inmates.[1]

This rapid increase in incarceration rates and corrections expenditures has led

We thank David Card, Enrico Moretti, Jasjeet Sekhon, and Uri Simonsohn for their very helpful input and the Russell Sage Foundation for its generous support of this project.

[1] Over this period, nominal expenditures on corrections increased from approximately $9 billion in 1982 to $68 billion in 2006. Adjusted for inflation, this represents a threefold increase in expenditures on corrections.

[*Journal of Law and Economics*, vol. 55 (May 2012)]

many to ask whether, on the margin, the benefits of incarceration exceed the costs (for example, see DiIulio and Piehl 1991; Donohue and Siegelman 1998; Levitt 1996; Jacobson 2005). Presumably, the chief benefit of higher incarceration rates is the crime avoided via the incapacitation of the criminally active and the crime prevented via the general deterrence of the potentially criminally active. The larger such incapacitation and deterrence effects are, the more likely that the value of incarcerating one more offender exceeds the explicit outlays and the more difficult-to-measure social costs of incarceration.

However, there is considerable disagreement about the size of such effects and, for most recent offenders, whether incapacitation and deterrence effects exist at all. Those who argue for small crime-incarceration effects note the lack of a strong correlation between aggregate crime and incarceration rates (Jacobson 2005; Western 2006) as well as the likelihood that the crime-reducing effects of incarceration are likely to be declining as the prison population increases. Researchers finding larger effects (Levitt 1996) emphasize a fundamental identification problem that is likely to bias estimates of crime-prison effects toward zero—namely, changes in behavior that increase criminal activity will simultaneously increase incarceration rates.

In this paper, we present new evidence for the effect of aggregate changes in incarceration rates on changes in crime rates that accounts for the potential simultaneous relationship between incarceration and crime. Our principal innovation is that we develop an instrument for future changes in incarceration rates based on the theoretically predicted dynamic adjustment path of the aggregate incarceration rate in response to a shock (from whatever source) to prison entrance or exit transition probabilities. Given that incarceration rates adjust to permanent changes in behavior with a dynamic lag (given that only a fraction of offenders are apprehended in any one period), one can identify variation in incarceration rates that is not contaminated by contemporary changes in criminal behavior. We isolate this variation and use it to tease out the causal effect of incarceration on crime rates.

We present a simple model of incarceration and crime in which steady-state incarceration rates are determined by the transition probabilities between the incarcerated and nonincarcerated populations. We use this model to derive a prediction regarding the lead 1-period change in incarceration rates based on the current disparity between the actual incarceration rate and the steady-state incarceration rate implied by the current-period transition probabilities describing movements into and out of prison. This predicted change serves as our instrument for actual future increases in incarceration. Absent controls, our instrument explains nearly one-fifth of the variance in 1-year changes in incarceration rates.

Using state-level data for the United States for the period 1978–2004, we find crime-prison elasticities that are considerably larger than those implied by ordinary least squares (OLS) estimates. For the entire period, we find average crime-prison effects with implied elasticities of between $-.06$ and $-.11$ for violent crime and between −.15 and −.21 for property crime. We also present results for two subperiods of our panel: 1978–90 and 1991–2004. Our instrumental variables (IV) estimates for the earlier period suggest much larger crime-prison effects, with elasticity estimates consistent with those presented by Levitt (1996), who analyzes a similar period yet with an entirely different identification strategy. For the later period, however, the effects of changes in prison population share on crime rates are much smaller. Our results indicate that recent increases in incarceration rates have generated much less value for investment in terms of crime rate reduction.

2. Incarceration and Crime: A Review of the Existing Research

Incarceration may affect the overall level of crime rates through a number of causal channels. To start, incarceration mechanically incapacitates the criminally active. In addition, the threat of incarceration may deter potential criminal offenders from committing a crime in the first place, a causal path referred to as general deterrence. Over the longer term, prior prison experience may either reduce criminal activity among former inmates who do not wish to return to prison (referred to as specific deterrence) or perhaps enhance criminality if a prior incarceration increases the relative returns to crime.

Criminological research on the effect of incarceration on crime rates has primarily focused on the potential incapacitation effects of prison. Much of this research indirectly estimates incapacitation effects by interviewing inmates regarding their criminal activity before their most recent arrest and then imputing from their retrospective responses the amount of crime that inmates would have committed. Results from this research vary considerably across studies (often by a factor of 10), a fact often attributable to a few respondents who report incredibly large amounts of criminal activity. The most careful reviews of this research suggest that, on average, each additional prison-year served results in 10–20 fewer serious felony offenses (see the discussion in Marvell and Moody [1994] and the extensive analysis of Spelman [1994, 2000]).[2]

By construction, these incapacitation studies provide only a partial estimate of the effect of incarceration on crime rates, since they are unable to detect

[2] More recent research in this vein attempts to directly estimate incapacitation effects by observing the criminal behavior of former inmates after release. Owens (2009) exploits a sentence disenhancement to estimate the effect of shorter sentences on overall crime. In 2003, the state of Maryland discontinued the practice of consideration of juvenile records in the determination of sentences for adult offenders between 23 and 25 years of age. Owens estimates that this change in sentencing procedures reduced the time served by 200–400 days for those adult offenders who had prior juvenile convictions. By observing arrests during the period when they would have been incarcerated had they been sentenced under the prior sentencing regime, Owens estimates that sentence disenhancement increases the number of serious offenses by roughly 2–3 index crimes per offender per year of street time.

whether potential offenders are deterred by the threat of incarceration.[3] Moreover, the likely unreliability of inmate self-reports and the large cross-study variation in results suggest the need for alternative strategies. Given these limitations, several scholars have attempted to estimate the overall effect of incarceration using aggregate crime and prison data. However, these studies must address an alternative methodological challenge—namely, the fact that unobserved determinants of crime are likely to create a simultaneous relationship between incarceration and crime.

Marvell and Moody (1994) are perhaps the first to estimate the overall incarceration effect using state-level panel regressions. The authors use a series of Granger causality tests and conclude that after first-differencing the data, within-state variation in incarceration is exogenous. Marvell and Moody subsequently estimate the effect of incarceration on crime using a first-difference model with an error correction component to account for the cointegration of the crime and prison time series. The authors estimate an overall crime-prison elasticity of −.16.

Levitt (1996) also estimates the effect of incarceration on crime using a state-level panel data model. Unlike Marvell and Moody, however, Levitt explicitly corrects for the potential endogeneity of variation in incarceration rates. Levitt exploits the fact that in years when states are under a court order to relieve prisoner overcrowding, state prison populations grow at a significantly slower rate than in states that are not under such court orders. Using a series of variables measuring the status of lawsuits for prisoner overcrowding as instruments for state-level incarceration rates, Levitt finds two-stage least squares estimates of crime-prison elasticities that are considerably larger than comparable OLS estimates, with a corrected property crime-prison elasticity of −.3 and a violent crime-prison elasticity of −.4.

To be sure, the relatively large estimates in Levitt (1996) have been criticized on the basis of the choice of instruments. For example, Western (2006) argues that most of the states being placed under court order were southern states during the 1980s and early 1990s. To the extent that Levitt is capturing a local

[3] There is a separate and growing literature that attempts to estimate the general deterrence effects of incarceration. Kessler and Levitt (1999) estimate the effect of sentence enhancements for violent crime on overall offending, arguing that the crimes receiving the enhancement would have resulted in incarceration regardless and that any short-term effect of the enhancement on crime is thus attributable to pure deterrence. Webster, Doob, and Zimring (2006), however, argue that the deterrence estimates in Kessler and Levitt are driven by crime rates that were already trending downward and thus are spurious. A separate set of studies attempts to estimate general deterrence effects by exploiting the discontinuous increase in sentences for offenses that occurred at 18 years of age. Levitt (1998) finds a decrease in offending when youth reach the age of majority, while Lee and McCrary (2005) find no evidence of such an effect. More recently, Drago, Galbiati, and Vertova (2009) exploit a unique feature of a 2006 Italian mass pardon to identify general deterrent effects. The Italian pardon released most inmates with 3 years or less remaining on their sentence. Those who reoffended after release faced an enhanced sentence through the addition of the remainder of their unserved time to whatever new sentence was meted out for the new postrelease offense. The authors find that those inmates who faced a longer sentence enhancement (conditional on observables) were less likely to reoffend after being released.

average treatment effect specific to the South during this period, the large estimates may not generalize to the country as a whole. Donohue and Siegelman (1998) argue that Levitt's chosen instruments may themselves be endogenous, as states that have had unusually large increases in prison populations are more likely to come under court order to relieve overcrowding.

In our assessment, Levitt is surely correct in arguing that OLS estimates of the prison-crime effect are likely to be biased toward zero by a reverse causal effect of crime on incarceration. Moreover, the first-stage relationship between incarceration and his lawsuit variables is strong, well documented, and well argued. Given the paucity of estimates that correct for the endogeneity of prison, however, as well as the quite large estimates presented by Levitt, further research on the bias in simple first-differenced or fixed-effect crime models is necessary. In what follows, we present an alternative identification strategy.

3. Methodological Framework

In this paper, we follow Marvell and Moody (1994) and Levitt (1996) in estimating the overall crime-prison effects using state-level panel data regressions. Our principal innovation is that we derive an instrument for future increases in incarceration rates on the basis of the predicted dynamic adjustment of incarceration rates to changes in the underlying transition probabilities describing the incarceration stochastic process. Among the benefits of our strategy, the principal one is that regardless of the source of the shock (for example, a change in underlying criminal behavior, increased enforcement, and longer sentences), the dynamic adjustment of the incarceration rate to any permanent shock provides exogenous variation in future incarceration changes that can be used to identify the crime-incarceration effect. A second benefit concerns the fact that the instrument can be defined for all periods and states, and thus we are able to explore whether the marginal incarceration effects are changing over time.

To illustrate our strategy, we first present a simple aggregate model of incarceration and crime in which we assume an exogenous shock to the underlying criminality of the populace and in which we assume further that criminal activity is not responsive to variation in the contemporaneous incarceration rate or enforcement (that is, there is no general deterrent effect). We subsequently discuss how the model is altered by the incorporation of behavioral responses to changes in incarceration rates.

3.1. A Simple Model of Incapacitation

Suppose that, at any given time, the members of a population can be described by the current state i, where $i = 1$ corresponds to not being incarcerated and $i = 2$ corresponds to being incarcerated. Define the vector $\boldsymbol{S}_t' = [S_{1,t}\ S_{2,t}]$, where $S_{i,t}$ is the proportion of the population in state i in period t. Assume that the periodic probability that any individual commits a crime is given by the constant parameter c. Assume further that only the nonincarcerated can commit crime—that is, incarceration mechanically incapacitates potential offenders. We also assume that the likelihood of being caught and sent to prison, conditional on committing a crime, is given by the parameter p. Taken together, these assumptions indicate that the probability of transition from nonincarceration to incarceration is simply cp, while the fraction flowing into prison between periods t and $t+1$ is given by $cpS_{1,t}$. Finally, we assume that the periodic probability of being released from prison is given by the parameter θ for all inmates and periods.

For any period t, the population distribution across the two states is determined by an equation relating current state population shares to last period's population shares

$$\boldsymbol{S}_t' = \boldsymbol{S}_{t-1}'\boldsymbol{T}, \tag{1}$$

where $\boldsymbol{T} = [T_{ij}]$ is a transition probability matrix with each element representing the likelihood that a person in state i transitions to state j. Given our assumptions, the transition matrix is given by

$$\boldsymbol{T} = \begin{bmatrix} 1 - cp & cp \\ \theta & 1 - \theta \end{bmatrix}, \tag{2}$$

where the first row of the matrix provides the survival and hazard functions (respectively) for the nonincarcerated and the second row provides the hazard and survival functions for the incarcerated.

Equation (1) gives the relationship between the distribution of persons across states in period t and the comparable distribution in period $t-1$ and suggests that this distribution changes over time according to the elements of $\boldsymbol{T}$. Given enough periods, however, the proportions not incarcerated and incarcerated will eventually settle to steady-state values. Assuming stability in the elements of $\boldsymbol{T}$, the steady state is defined by the equation

$$\boldsymbol{S}^{*\prime} = S^{*\prime}\boldsymbol{T}, \tag{3}$$

where we have dropped the time subscript and added an asterisk to indicate the steady-state population share vector. When combined with the constraint $S_1^* + S_2^* = 1$, the steady-state population shares can be expressed as

$$S_1^* = \frac{\theta}{cp + \theta}, \qquad S_2^* = \frac{cp}{cp + \theta}. \tag{4}$$

With these specific values, the equilibrium crime rate can be derived by multiplying the proportion not incarcerated by the probability that someone commits a crime. This yields

$$\text{Crime}^* = cS_1^* = c(1 - S_2^*) = \frac{c\theta}{cp + \theta}. \tag{5}$$

A comparative static analysis of the steady-state shares in equation (4) and the equilibrium crime rate in equation (5) can be used to highlight the fundamental identification problem faced by empirical studies of the crime-incarceration effect that make use of aggregate data. Within the context of the simple mechanical model derived here, such studies seek to uncover the individual propensity to commit crime—that is, the parameter c. Multiplying -1 by this parameter gives the reduction in crime that would occur through increased incapacitation with a one-person increase in the incarceration rate. The aggregate analysis seeks to uncover this parameter by empirically estimating the effect of a change in the incarceration rate on crime rates, or $\partial \text{Crime}^*/\partial S_2^*$. Whether this aggregate relationship reveals the parameter c will depend on which of the three underlying parameters is driving the change in incarceration and crime.

For example, suppose that a sentence enhancement reduces the value of the parameter θ (effectively increasing sentence length). In this instance, the empirically observed change in crime co-occurring with the observed change in incarceration would be given by $\partial \text{Crime}^*/\partial S_2^* = (\partial \text{Crime}^*/\partial \theta) / (\partial S_2^*/\partial \theta) = -c$. Because this is the parameter we seek to estimate, the aggregate analysis in this instance would yield an unbiased estimate. Similarly, when variation in incarceration and crime rates is driven by an exogenous shock to the apprehension parameter p, the empirical estimate of the change in the crime rate caused by a change in the incarceration rate will also equal $-c$. That is to say, exogenous policy variation in the incarceration rate identifies the parameter of interest and reveals the crime rate reduction caused by an increase in the incarceration rate.

However, a change in incarceration rates caused by an exogenous change in criminal behavior will not uncover the criminality parameter. An increase in criminal behavior (operationalized as an increase in the parameter c) will cause both an increase in incarceration rates and an increase in crime rates—that is, $\partial S_2^*/\partial c$, $\partial \text{Crime}^*/\partial c > 0$. The ratio of the latter derivative to the first (corresponding to the naive empirical estimate of $\partial \text{Crime}^*/\partial S_2^*$) yields the solution θ/p. Because both of these parameters are positive, variation in incarceration and crime rates caused by exogenous shocks to criminal behavior may create the false impression that higher incarceration rates lead to higher crime rates. At a minimum, the dependence of crime and incarceration rates on underlying variation in criminal propensities will positively bias empirical estimates of $\partial \text{Crime}^*/\partial S_2^*$ that do not account for this simultaneity problem.

Most panel data studies of crime and incarceration do not regress changes in steady-state crime rates on changes in steady-state incarceration rates, since shocks to the underlying parameters of the two variables will induce multiperiod adjustment processes toward new steady-state values. To the extent that the observed temporal variation in crime and incarceration is at a relatively high frequency (annual data, for instance), changes in these variables will reflect both responses to contemporaneous changes in underlying transition parameters and the dynamic adjustment between equilibria caused by past changes in the underlying parameters. The innovation that we highlight here exploits variation in incarceration rates caused by past changes in the various transition probabilities that is arguably independent of contemporaneous and subsequent changes in parameter values. As we show below, variation in incarceration rates associated with such longer term adjustment to earlier shocks is plausibly exogenous and can be used to identify the crime-prison effect.

To illustrate this point, we attribute the adjustment paths of crime and incarceration to a permanent change in the propensity to commit crime and show that variation along these adjustment paths beyond the first-period response can be used to identify the parameter of interest c. As the preceding comparative static analysis demonstrated, this is the worse-case-scenario source of variation for the purpose of estimating the crime-prison elasticity.[4] With dynamically lagged responses of incarceration to this change, however, suitable variation in incarceration rates can be isolated to measure the underlying causal relationship.

Suppose the system is initially in steady state, with a value for the criminality parameter equal to c_0 at time $t = 0$. The propensity to commit crime then increases at $t = 1$ from c_0 to c_1. For any period $t > 0$, the proportion incarcerated is given by

$$S_{2,t} = S_{1,t-1}c_1p + S_{2,t-1}(1 - \theta), \tag{6}$$

where we have now reintroduced the time subscript since it is no longer presumed that we are in the steady state at any point in time. Substituting $1 - S_{2,t\text{-}1}$ for the share not incarcerated in period $t - 1$ and then rearranging yields the expression

$$S_{2,t} + S_{2,t-1}(c_1p + \theta - 1) = c_1p, \tag{7}$$

which is in the form of a simple linear difference equation. To derive an explicit description of the dynamic path of incarceration as a function of time and the underlying transition parameters, we need to define the initial condition at $t = 0$ for the incarceration rate. Since the system was in steady state before the shock, the incarceration rate at time $t = 0$ is $S^{\star}_{2,t=0} = c_0p/(c_0p + \theta)$. With this initial condition, solving equation (7) as a function of the parameters and time gives the expression

$$S_{2,t} = \left(\frac{c_0p}{c_0p + \theta} - \frac{c_1p}{c_1p + \theta}\right)(1 - c_1p - \theta)^t + \frac{c_1p}{c_1p + \theta},$$

which can be rewritten as

$$S_{2,t} = (S^{\star}_{2,t=0} - S^{\star}_{2,t>0})(1 - cp - \theta)^t + S^{\star}_{2,t>0}, \tag{8}$$

[4] Similar arguments apply to permanent changes in the composite probability of apprehension and conviction or the probability of release from prison. In fact, the instruments in Levitt (1996) are most likely operating through an exogenous shift in θ. Here we focus on the adjustment to changes in the criminality parameter c, since it is generally thought to be the most likely contaminating factor that was omitted. Nonetheless, the ensuing argument and instrumental variables strategy applies to permanent changes in any of the parameters of the transition probability matrix.

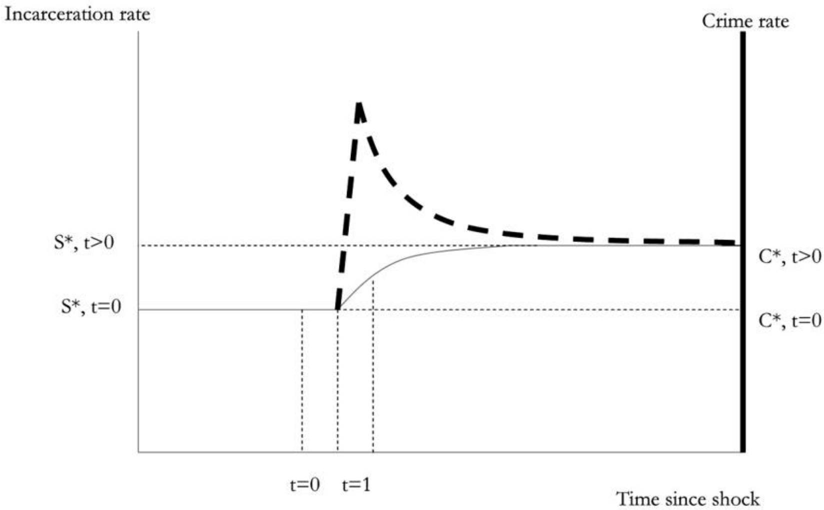

Figure 1. The dynamic adjustment path of incarceration and crime rates

where $S^{*}_{2,t=0} = c_0 p/(c_0 p + \theta)$ is the old steady-state incarceration rate prior to the change in criminality and $S^{*}_{2,t>0} = c_1 p/(c_1 p + \theta)$ is the new steady-state incarceration rate that will eventually be reached given stability in the parameters and enough time.

Equation (8) shows that the incarceration rate at any time $t > 0$ is equal to the new steady-state incarceration rate (the second term on the right) plus a proportion of the disparity between the old and new steady-state rates. Since the first term is negative, and since $0 < (1 - cp - \theta) < 1$ for typical values of these parameters,[5] equation (8) depicts a stable process whereby the incarceration rate approaches the new steady state from below. The adjustment path is depicted in Figure 1. Note that the incarceration rate increases between $t = 0$ and $t = 1$ because of the increase in the criminality parameter. Subsequent increases, however, are not driven by further changes in c, since we have assumed a one-time permanent shock to criminality. Rather, subsequent increases reflect the dynamic multiperiod adjustment of incarceration toward its new equilibrium rate. For annual data for U.S. states, a typical value for θ is roughly .5, while a

[5] In theory, it is plausible that the term $(1 - cp - \theta)$ could fall below zero. This would require a very high release probability and a very large inflow rate into prison. If this were the case and if $-1 < 1 - cp - \theta < 0$, then the incarceration rate would still converge asymptotically to the higher steady-state value. However, rather than approach the steady state from below, the adjustment path would oscillate above and below the long-run steady state, with the oscillation variance diminishing with time. In practice, the prison release rate in the United States hovers around .5, and the transition into prison, cp, is consistently below .01. Hence, for practical purposes it is safe to assume that $0 < 1 - cp - \theta < 1$.

typical value for cp (the flow rate into prison) is, at most, .01. These values, combined with equation (8), suggest that the incarceration rate becomes quite close to its new equilibrium value after 5 or 6 years.

We can derive a similar adjustment path for the crime rate. Substituting the time path for incarceration into equation (5) and rearranging yields the expression

$$\text{Crime}_t = c_t(S^{*}_{2,t>0} - S^{*}_{2,t=0})(1 - cp - \theta)^t + c_t(1 - S^{*}_{2,t>0}), \quad (9)$$

where the crime rate at time $t > 0$ consists of two components: the new steady-state crime rate (the second terms on the right-hand side of equation [9]), and the deviation from the new steady state associated with the dynamics adjustment (the first term). Here the adjustment term is positive and approaches zero as t increases, which implies that the crime rate approaches its new equilibrium from above. Given that the new steady-state crime rate will exceed the old steady-state crime rate, equation (9) indicates that in response to a permanent increase in criminality, the crime rate increases discretely and then declines to the new equilibrium over time. The time path for this variable is also depicted in Figure 1.

The patterns observed in Figure 1 hint at our identification strategy. Between periods $t = 0$ and $t = 1$, both crime and incarceration rates increase as a result of the discrete increase in the criminality parameter from c_0 to c_1. Clearly, the positive covariance between the two variables for this first difference (both crime and incarceration rates increase) is driven by the change in criminality. Thus, a regression of a series of such first-period changes in the crime rate against a first-period change in the incarceration rate will yield a spurious positive coefficient.

This is not the case, however, for subsequent changes in these series. For all changes beyond the first, the criminality parameter is held constant, yet the incarceration rate increases as it approaches its new equilibrium rate. With regard to the crime rate, following the initial discrete increase, subsequent increases in the incarceration rate decrease the crime rate by incapacitating a greater proportion of the population. In other words, the decline in the crime rate along the adjustment path beyond the change between periods 0 and 1 is driven by the increase in the incarceration rate. In fact, the ratio of the changes in the crime rate to the change in the incarceration rate for any of these subsequent periods would yield -1 times the criminality parameter—that is, the incapacitation effect that we seek to estimate. Thus, if one could discard the variation associated with the initial shock and isolate variation in subsequent movements associated with the dynamics adjustment to the shock, one could identify the incapacitation effect of marginal increases in the incarceration rate.

The exogeneity of subsequent changes in the incarceration rate in this model is best illustrated by deriving explicit expressions for the change in incarceration and crime rates following permanent shocks to criminality. Let $\Delta S_{2,t} = S_{2,t+1} - S_{2,t}$ and $\Delta C_t = C_{t+1} - C_t$. From equation (8), explicit expressions for 1-period changes in incarceration rates for $t = 0$, $t = 1$, and $t > 1$ are

$$\Delta S_{2,0} = (S^{*}_{2,t>0} - S^{*}_{2,t=0})(c_1 p + \theta),$$
$$\Delta S_{2,1} = (S^{*}_{2,t>0} - S^{*}_{2,t=0})(c_1 p + \theta)(1 - c_1 p - \theta), \tag{10}$$
$$\Delta S_{2,t} = (S^{*}_{2,t>0} - S^{*}_{2,t=0})(c_1 p + \theta)(1 - c_1 p - \theta)^t.$$

Given that $(1 - c_1 p - \theta) < 1$, the first change in the incarceration rate is the largest, and each subsequent change diminishes in size.

By equation (9), an explicit expression for the first-period change in the crime rate is

$$\Delta \text{Crime}_o = -c_1 \Delta S_{2,0} + (c_1 - c_0)(1 - S^{*}_{2,t=0}), \tag{11}$$

which has two components, one of which we are interested in uncovering. The second term on the right-hand side of equation (11) gives the change in the crime rate associated with the increase in criminality, holding the incarceration rate at its equilibrium in period $t = 0$. This component is positive and drives the initial spike in crime rates. The first term on the right-hand side of equation (11) shows the association between the decline in crime rates and the first-period increase in incarceration rates. Thus, the discrete increase in crime rates in Figure 1 entails the sum of two effects: the effect of an increase in criminality (the larger of the two) and the partially offsetting effect of the contemporaneous increase in incarceration.

In practice, we observe the total change in the crime rate and the change in the incarceration rate, and we wish to estimate the coefficient $-c_1$ associated with the first component in equation (11). We do not observe the second term on the right-hand side of equation (11), and thus it is swept into the residual of an OLS regression. Given that the contemporaneous change in criminality will be positively correlated with the contemporaneous change in the incarceration rate, an OLS regression of ΔCrime_0 on $\Delta S_{2,0}$ will yield a positively biased estimate of $-c_1$. This argument is similar to the identification problem that we highlighted in the comparative static analysis.

However, changes subsequent to the first-period change will not suffer from this bias. To see this, the explicit expression for the next change in crime is given by

$$\Delta \text{Crime}_1 = -c_1 \Delta S_{2,1}. \tag{12}$$

Here the crime rate change is a function of the change in incarceration rates alone. This follows from the fact that we are modeling a one-time permanent increase in criminality, and thus the contaminating second term in equation (11) drops out for all subsequent changes in crime rates until the crime rate reaches its new steady-state level. Most important, taking the ratio of equation (12) to the second line of equation (10) yields the parameter of interest $-c_1$.

Together, equations (10) and (12) provide the heart of our identification strategy. The second line of equation (10) provides a prediction for the change in incarceration rates between periods 1 and 2 associated with an increase in criminality between periods 0 and 1. Since this predicted increase is not driven by contemporaneous changes in criminal behavior between periods 1 and 2, a variable constructed from the second line of equation (10) could serve as an instrument for actual changes in incarceration rates. Our principal strategy is to use this prediction to instrument changes in the incarceration rate in a series of crime models where the incarceration rate is the principal explanatory variable of interest. In other words, we estimate the difference equation (12) where the actual change in the incarceration rate is instrumented with the corresponding predicted change in the incarceration rate from equation (10). In essence, the strategy identifies variation in that incarceration rate that would have occurred anyway and uses this variation to identify various crime-incarceration effects.

Although we have not performed a similar analysis of the dynamic adjustment processes of crime and incarceration rates to underlying shocks to the policy parameters p and θ, it should be noted that, to the extent that such shocks are causing variation in crime and incarceration rates, a simple OLS regression of crime rates on incarceration rates is indeed identified. To be specific, making reference to the change in the crime rate in equation (11), the contaminating second term occurs because of the exogenous shock to criminality. If changes in incarceration and crime rates are being driven by shocks to parameters other than the criminality parameter, this contaminating factor drops out of the equation. Hence, the crime rate responds to a change in p or θ only through the impact of these parameters on the incarceration rate, and thus the OLS regression provides an unbiased estimate of the pure incapacitation effect.

3.2. *Behavioral Responses to Changes in Criminality and Enforcement*

Thus far, we have assumed that the underlying transition parameters in our aggregate model of crime and incarceration do not respond to one another either instantaneously or with a dynamic lag. Clearly, this assumption is unrealistic, as we would expect policy makers to respond to changes in the prevalence of criminal behavior and potential criminals to perhaps respond to changes in policy. For our narrow purposes, we are particularly interested in whether incorporating behavioral responses compromises our identification strategy.

In the Appendix, we provide a detailed discussion that sequentially relaxes some of the behavioral assumptions that we have implicitly made. Here we highlight how the incorporation of such behavioral responses is likely to impact our first-stage prediction, the potential exogeneity of our instrument, and the interpretation of our empirical results. Although we work through quite specific responses and the dynamic structure below and in the Appendix, the following general points can be gleaned from this exercise.

First, allowing for endogenous responses of the policy parameters to an underlying criminality shock will influence the strength of our first-stage prediction (two-stage least squares model) but does not compromise the exogeneity of our instrument. For several possible policy responses, we show that the lead 1-period change in incarceration rates can be decomposed into a component equal to our prediction, assuming no behavioral response (via equation [10]) as well as additional components associated with the second-order responsive changes in p and θ. Since these parameters influence crime only through the incapacitation and deterrence effects of incarceration, we can still identify a causal effect as long as our instrument predicts significant variation in the change in incarceration rates.

Second, when we allow for a reciprocally responsive relationship between criminality and enforcement, we can no longer interpret the key coefficient as measuring a pure incapacitation effect (as in our simple model above). To be specific, increases in enforcement caused by an exogenous shock to behavior may induce any increase in criminality to be subsequently dulled by the higher incarceration risk. When this is the case, the empirical association between incarceration and crime will be driven by incapacitation as well as deterrence. In the context of our instrumental variables strategy, our estimates will yield a biased estimate of the pure incapacitation effect. However, our estimate of the total effect of prison on crime is unbiased as long as we interpret the estimate more broadly.

Finally, there are certain conditions under which the identification strategy proposed here would fail. The most obvious would be when changes in the criminality parameter exhibit negative serial correlation for reasons that are independent of policy. For example, if a wave of drug-related crime today leads to the emergence of informal social controls that are external to the criminal justice system that subsequently reduces criminality, the predicted future increase in incarceration rates may be spuriously inversely related to future decreases in crime rates.

In the Appendix, we provide a more detailed analysis of the consequences of allowing for the following behavioral relationships.

Allowing Enforcement to Respond to Changes in Criminality: $p = p(c)$. One might hypothesize that p may be either increasing or decreasing in the degree of criminality. If policy makers increase enforcement in response to an increase in c, an elevated propensity to commit crime may be matched by an elevated incarceration risk. On the other hand, an increase in c may dilute enforcement resources and reduce the risk of incarceration. Such a policy reaction should not compromise the exogeneity of our instrument, although the timing of the response may impact the strength of our first-stage prediction. If p responds to changes in c instantaneously, the behavioral response will either speed up or slow down the adjustment processes of crime and incarceration rates to their new equilibrium values (depending on the sign of dp/dc). If p responds with a lag, our instrument will either overpredict or underpredict the actual change in incarceration rates. However, the proposed instrument is still orthogonal to the second-stage error term.

Allowing Sentence Severity to Respond to Changes in Criminality: $\theta = \theta(c)$. Assessing the effect of a change in sentence length on our identification strategy by necessity requires a dynamic analysis, since a change in sentence length today will not impact incarceration rates until today's cohort of admitted inmates reach their counterfactual release dates under the prior sentencing regime. Assuming a 1-period lag in the response of θ to a change in c, an increase in sentence length (operationalized as a decrease in θ) in response to an increase in c implies that our instrument will underpredict the change in incarceration rates after the initial behavioral shock. Although this introduces error into our first-stage prediction, the instrument is still exogenous to the unobserved determinants of future changes in crime rates.

Allowing Criminality and Enforcement to React Reciprocally: $p = p(c)$, $c = c(p)$. The implications of allowing simultaneous determination of the criminality and incarceration risk parameters for our identification strategy will depend on whether these adjustments will occur instantaneously or over time. Moreover, if the reaction processes are dynamic, the timing and sequencing of the reactions are important in assessing how such behavioral responses would impact the interpretation of our empirical results. If we assume that criminality responds to changes in p instantaneously, while enforcement responds to changes in c with a lag, our proposed instrumental variables strategy will yield a biased estimate of a pure incapacitation effect. This bias is driven by the fact that the error term in the second-stage equation relating changes in crime rates to changes in incarceration rates will include a component reflecting the behavioral response of criminal behavior to enhanced enforcement (a term that will be negatively correlated with our predicted change in incarceration). However, since this component is essentially a general deterrent effect, the structural estimate of the effect of incarceration on crime still represents a causal effect, as long as this estimate is interpreted as the overall impact of a change in incarceration rates (incapacitation plus deterrence). With regard to the first-stage prediction, reciprocal reactionary responses between p and c imply that future increases in incarceration rates in response to a change in c may be either smaller or larger than a nonbehavioral model would predict, since the effect of enhanced enforcement on incarceration rates is offset by subsequent deterrence-induced declines in criminality.

Allowing Future Change in Criminality to Respond to a Previous Change in Criminality. Suppose that current increases in criminal behavior cause subsequent decreases in criminality because of a revulsion on the part of those likely to commit crime in response to the consequences of an initial crime spike. Such an effect would induce negative serial correlation in changes in the criminality parameter and would likely induce a spurious negative correlation between our instrument (which is increasing in the past period's increase in criminality behavior) and future changes in crime rates (which would be negatively impacted by the hypothesized reactive behavior). Of course, if the periodicity of our data is such that the 1-period lead prediction we employ as an instrument predicts changes in incarceration before such revulsion wells up and impacts crime, our IV strategy would still be valid.

To summarize, with the exception of the final possibility, incorporating behavior into our model may impact the precision of our first-stage prediction but does not compromise the exogeneity of the proposed instrument. If we allow criminal behavior and corrections policy to respond simultaneously to one another, we need to interpret the IV results as an overall effect of incarceration on crime rather than as an estimate of a pure incapacitation effect. Nonetheless, the IV estimates still carry a causal interpretation.

Our strategy would not be suitable if criminal behavior reacts negatively to previous increases in criminal behavior (through channels not mediated by a change in enforcement or sentencing). Above we offer the example of the emergence of informal social controls intended to mitigate an increase in criminality. However, there are strong reasons to believe that such responses would likely take more than 1 year (the time frame of our prediction) and thus are unlikely to compromise our results. Nonetheless, we acknowledge this potential weakness.

3.3. Heterogeneity in the Propensity to Commit Crime

In our theoretical model, we made the simplifying assumption that the criminality parameter c was constant across the population. This assumption thus permits identification of a constant incapacitation effect. To be sure, this parameter most likely has a nondegenerate distribution among the general public, and the conditional expectation of c among the incarcerated (as well the value on the margin for recent prison entrant) likely depends on the overall incarceration rate (with all else held equal, one would expect lower average values of c among inmates the higher the incarceration rate). With heterogeneity in the criminality parameter, estimates of the joint incapacitation/deterrence effects should be interpreted as local average treatment effects specific to the period covered by the underlying data. One of the key motivations driving this analysis is to test for such heterogeneity, specifically with regard to alternative periods with different average incarceration rates.

4. Data Description and Documentation of the First-Stage Relationship

Implementing our identification strategy requires that we obtain information on the transition probabilities between incarceration and nonincarceration by state and year. Our strategy also requires that we identify permanent changes in underlying transition probabilities. Finally, there are very few states and periods in which changes in incarceration rates fit the model of a one-time increase in criminality with a delayed dynamic adjustment. In fact, over the past 20 or more years, most states have experienced repeated increases in prison admission rates. Thus, we must adapt our strategy to incorporate these serial shocks. Here we describe the data for this project and the manner in which we use these data to implement our identification strategy.

Our first task is to estimate the transition probabilities by state and year, since

Table 1
Illustration of the Calculation of the Predicted Change in Incarceration Rates for New York, 1979–82

	1979	1980	1981	1982
Current incarceration rate ($S_{2,t}$)	118.39	125.33	147.30	161.39
Admission rate (cp)		.00059	.00071	.00072
Release rate (θ)		.432	.329	.360
Equilibrium incarceration rate based on current transition probabilities ($S^*_{2,t>0} = [cp/(cp + \theta)] \times 100{,}000$)		135.87	215.61	199.97
Incarceration rate at $t = 0$ ($S_{2,0}$)		118.39	125.33	147.30
Predicted change in incarceration rate[a] $[(S^*_{2,t>0} - S_{2,0}) \times (1 - cp - \theta)(cp + \theta)]$		4.29	19.94	12.15
Actual change in incarceration rate[a]		21.97	14.09	13.64

[a] From $t = 1$ to $t = 2$.

our proposed instrument requires information about cp and θ. Doing so requires four pieces of information: aggregate annual flows into prison, aggregate annual flows out of prison, the stock of prisoners for a given year, and an estimate of the total state population. We obtain data on aggregate flows into and out of prison by state and year from the National Prison Statistics database.[6] These data provide the total number of admissions and total number of releases from prison within a calendar year. Data on the stock of prison inmates under each state's jurisdiction come from the Bureau of Justice Statistics and measure the stock of inmates as of December 31 of the stated calendar year. State-level population data are from the U.S. Census Bureau. We estimate the probability of transition from nonincarceration to incarceration by dividing the total number of prison admissions by the total state population less the inmate population. We calculate the incarceration-nonincarceration transition probability by dividing the total annual number of prison releases by the stock of prison inmates.

The next task involves adapting our IV regression to the fact that states are subject to serial shocks in transition probabilities rather than single shocks. Our manner of doing so is illustrated for 3 years for the state of New York in Table 1. Note that in each year, the equilibrium value exceeds the actual incarceration rate. For any given year, we designate $t = 0$ as the previous year. For example, for the purpose of predicting the increase in incarceration rates between 1980 and 1981, we designate 1980 as $t = 1$ and 1979 as $t = 0$. In predicting the change between 1981 and 1982, we set $t = 1$ for 1981 and $t = 0$ for 1980, and so on. Thus, the starting value for the dynamic adjustment for any given year is always defined as the 1-period lagged incarceration rate.

The predicted change in the incarceration rate between $t = 1$ and $t = 2$ is shown for each year on the basis of all of the values that are already determined

[6] For data from the National Prisoner Statistics database on year-end totals, admissions, and releases by state, see Bureau of Justice Statistics, Spreadsheets (http://bjs.ojp.usdoj.gov/content/dtdata.cfm). All data used in this study are available from the authors on request.

by time period $t = 1$ (with the exception of the initial incarceration rate, which is determined by $t = 0$), using the second expression in equation (10). Actual change in incarceration rates between $t = 1$ and $t = 2$ is calculated by using the predicted change in incarceration rates as an instrument for the first differences in incarceration rates.

Finally, our identification strategy requires that we identify permanent changes to the underlying transition probabilities (which may subsequently be enhanced or diminished by future changes in these probabilities). To the extent that the observed changes in empirical admission and release hazards reflect temporary rather than permanent changes, our instrument will serve as a poor predictor of future actual change in incarceration rates.[7] To minimize the influence of temporary shocks to the transition probabilities, we first smooth the transition probability time series for each state and use the smoothed series to construct the predicted change in incarceration rates as illustrated in Table 1. For each state, we estimate a simple regression where a given transition probability is regressed on an eighth-order polynomial time trend. We then calculate the predicted value for the transition probability from the estimated regression function. We estimate this model for each of the 50 states plus Washington, D.C., for the admission hazard as well as the release hazard (102 models in all). These predicted transition probabilities are then used to construct our instrumental variable.[8]

Our panel data set covers the period 1978–2004 and the 50 states and Washington, D.C.[9] Figure 2 presents a simple bivariate scatterplot (weighted by state-level population counts) of the actual annual changes in incarceration rates for the entire panel against the predicted changes from equation (10) (using the smoothed state-level transition probabilities). Several notable patterns stand out. First, there is a strong correlation between our instrument and actual changes in incarceration rates (approximately .40), with the instrument explaining roughly 16 percent of the variation in annual changes. Second, the lion's share of predicted increases in incarceration are positive (approximately 86 percent), which suggests that for most states and periods, the observed incarceration rate is below the steady-state rate implied by the value of their transition probabilities. Finally, the coefficient on the instrument is substantially less than 1 (.61), which suggests that the instrument is overpredicting the actual change in incarceration rates. Note that this is consistent with a reciprocal responsiveness between criminal propensity and enforcement, with subsequent declines in criminality occurring in response to enhanced enforcement moderating the impact of a crime

[7] Moreover, if temporary increases in criminality cause subsequent increases in incarceration rates because of time lags between arrest and incarceration or a spurt of criminal activity at the end of the year, temporary shocks to the transition probabilities may induce a spurious negative relationship between our instrument and crime.

[8] In the remainder of Section 4, we discuss results that use the unsmoothed parameter values to construct the instrument as well.

[9] There are a few missing observations for Alaska and two missing observations at the end of the time series for Washington, D.C., when the metropolitan area abandoned its prison system.

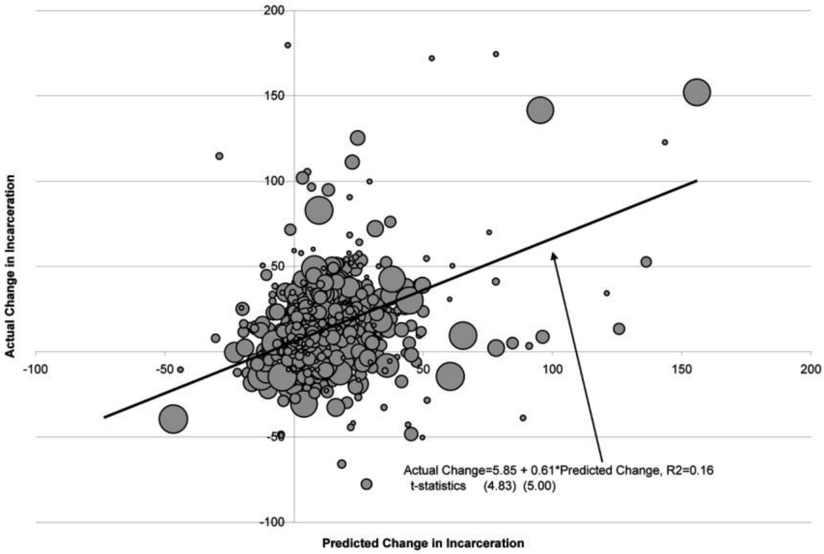

Figure 2. Actual annual versus predicted changes in state-level incarceration rates

shock on incarceration rates (see the Appendix for a detailed discussion of this possibility).

Table 2 assesses the robustness of this first-stage relationship to the inclusion of year fixed effects, state fixed effects, and a series of state-level controls for changes in the age structure, the percentage of minority residents, the percentage of poor residents, the state unemployment rate, and state per capita income.[10] All regressions are weighted by state-level population. We tabulate and report robust standard errors that permit clustering by state in the underlying error variance-covariance matrix. Adding year effects removes the influence of any factor affecting incarceration over time that is constant across states. Since the dependent variable (as well as all of the explanatory variables) is specified in first differences, adding a complete set of state fixed effects adjusts for state-specific linear time trends in incarceration rates after adjusting for common national year-to-year changes. Column 1 repeats the simple bivariate regression depicted in Figure 2, and regression (2) adds observable covariates. Regression

[10] Data on the percentage of state residents who are within given age groups and who are black are from U.S. Census Bureau, Population Estimates (http://www.census.gov/popest/data/historical/index.html). Data on poverty rates by state are from U.S. Census Bureau, State and County Interactive Tables (http://www.census.gov/did/www/saipe/county.html). Data on per capita income by state are from Bureau of Economic Analysis, Regional Economic Accounts (http://www.bea.gov/regional/). Finally, data on state unemployment rates are from U.S. Bureau of Labor Statistics, Databases, Tables and Calculators by Subject: Local Area Unemployment Statistics (http://www.bls.gov/data/#unemployment). All of the data used in the project are available from the authors on request.

Table 2

First-Stage Effect of the Predicted Change in Incarceration Rates (Based on Last-Period Shock) on the Current Change in Incarceration Rates

	(1)	(2)	(3)	(4)
Predicted ΔIncarceration Rate	.605 (.121)	.603 (.133)	.549 (.143)	.513 (.167)
ΔPopulation, by age (%):				
0–17		−2.415 (2.715)	−.684 (2.530)	−2.236 (3.057)
18–24		−3.871 (4.989)	.393 (3.483)	−1.109 (4.009)
25–44		−2.602 (3.492)	−1.074 (4.565)	−1.968 (5.236)
45–64		−3.619 (4.472)	1.752 (5.137)	2.386 (5.438)
ΔUnemployment Rate		−1.595 (.659)	−1.430 (.861)	−1.378 (.902)
ΔPoverty Rate		−.943 (.532)	.078 (.547)	.091 (.552)
ΔBlack (%)		.819 (.135)	.736 (.203)	.793 (.203)
ΔPer Capita Income		−.008 (.002)	−.005 (.002)	−.004 (.002)
Year effects	No	No	Yes	Yes
State effects	No	No	No	Yes
R^2	.164	.190	.312	.328
F-statistic (*p* value)	24.96 (<.0001)	20.63 (<.0001)	14.83 (<.0003)	9.45 (<.0035)

Note. The dependent variable is ΔIncarceration Rate. Robust standard errors, clustered by state, are in parentheses. All models include a constant terms and are weighted by the state-year populations. The *F*-statistic is from a test of the significance of the instrumental variable. $N = 1{,}321$.

(3) adds year effects to the specification in regression (2), while regression (4) adds state effects. While the coefficient on the predicted change in incarceration diminishes slightly across specifications and the standard error on the point estimate increases, the instrument is highly significant in all specifications. The *F*-statistics from a simple test of the significance of the instrument are greater than 10 in the first three specifications and fall slightly short of 10 in the final specification. Thus, the first-stage relationship is quite strong for the overall panel and persists with the inclusion of observable covariates.

We use the first-stage models in Table 2 to identify the effect of changes in incarceration on changes in crime rates. With regard to our dependent variables, we test for effects on all of the seven part I felony offenses included in the Federal Bureau of Investigation's Uniform Crime Reports (UCR). The UCR count all serious felony offenses reported to the police by state and year. Average crime rates for individual offenses and overall violent crime and overall property crime are presented in Table 3, along with the average state-level incarceration rate. Table 3 also presents descriptive statistics for two subperiods (1978–90 and 1991–2004), which we later use to stratify the period to estimate period-specific incarceration effects.

The means noted by subperiod suggest that the overall violent crime rate increased slightly, while overall property crime declined considerably in the latter period relative to the former. Data from victimization surveys indicate that both property and violent crime rates declined substantially over these periods. The higher violent crime rates in the latter period most likely reflect increased re-

Table 3
Descriptive Statistics for Crime and Incarceration Rates for the Overall Sample Period and Subperiods

	Average	SD	Within-State SD
1978–2004:			
Violent crime	596.35	256.50	124.70
Murder	7.85	4.05	2.34
Rape	36.03	11.69	6.63
Robbery	205.52	127.34	66.32
Assault	346.95	151.49	74.10
Property crime	4,503.20	1,189.87	817.27
Burglary	1,120.32	438.50	356.33
Larceny	2,875.62	701.85	431.54
Motor vehicle theft	507.27	223.40	138.55
Incarceration rate	302.62	163.70	132.10
1978–90:			
Violent crime	594.19	264.07	76.26
Murder	8.85	3.99	1.54
Rape	36.33	12.24	5.01
Robbery	226.38	145.17	36.07
Assault	322.63	138.59	58.28
Property crime	4,929.62	1,191.04	454.14
Burglary	1,392.10	423.60	217.76
Larceny	3,024.91	707.41	260.95
Motor vehicle theft	512.62	233.75	103.13
Incarceration rate	186.38	87.07	56.61
1991–2004:			
Violent crime	598.09	250.23	137.96
Murder	7.06	3.91	2.10
Rape	35.77	11.24	5.36
Robbery	188.71	108.04	67.44
Assault	366.53	158.43	72.95
Property crime	4,160.04	1,072.13	679.24
Burglary	901.59	308.77	205.33
Larceny	2,755.48	673.73	377.09
Motor vehicle theft	502.96	214.61	131.14
Incarceration rate	396.20	150.44	78.74

Note. Crime rates are number of incidents per 100,000 state residents; incarceration rates are number of inmates per 100,000 state residents. All values are weighted by the state-year population.

porting of violent crime to police,[11] a problem in the UCR data frequently noted in past research (Donohue and Siegleman 1998; Levitt 1996; Spelman 2000). In light of this problem, all of the crime models that we estimate below include year fixed effects.

[11] In fact, analysis by the Bureau of Justice Statistics shows that the crime-reporting rates have increased over the past 2–3 decades. See Bureau of Justice Statistics, Facts at a Glance (http://bjs.ojp.usdoj.gov/content/glance/tables/reportingtypetab.cfm).

5. Empirical Results Using the Entire Sample Period

Table 4 presents a series of regression model estimates in which the dependent variable is either the annual change in the overall violent crime rate or the annual change in the overall property crime rate. The first-stage models for the IV results correspond to specifications (3) and (4) in Table 2. Beginning with the results for violent crime, we find that in both OLS models the coefficient on the change in incarceration is small, positive, and statistically insignificant. Instrumenting with the predicted change in incarceration turns the coefficient negative, yet neither IV estimate is statistically significant. The magnitude of the estimate in the final specification suggests that each additional inmate reduces the annual number of violent crimes by .23 incident. This is a considerably larger effect than the small positive impact (.038) suggested by the corresponding OLS model; however, both point estimates have high variance.

Turning to the results for overall property crime, we find that there is more consistent evidence of a significant negative effect of incarceration rates on crime rates in all models. In the models that omit state fixed effects, a one-person increase in the incarceration rate is predicted to reduce the property crime rate by approximately 1. The IV estimate, however, indicates a much larger effect, with a coefficient on the change in incarceration rate of −2.315. Both estimates are statistically significant, with the IV estimate significant at the 1 percent level of confidence. Adding state fixed effects increases these estimates, marginally for the OLS model (to −1.109) but appreciably for the IV model (to −3.272). Similar to the results for violent crime, the juxtaposition of the OLS and IV estimation results strongly suggests that OLS estimates are biased toward zero by the simultaneous determination of incarceration and crime.

To facilitate comparison with previous research, Table 4 also presents the implied crime-prison elasticities for each specification.[12] For violent crime, conversion of our IV estimates to elasticities yields effects of −.06 to −.11. For property crime, the corresponding elasticity estimates range from −.15 to −.21. Levitt (1996) reports violent crime–prison elasticities between −.38 and −.42 and property crime–prison elasticities of −.26 to −.32. Marvell and Moody (1994) find a total crime–prison elasticity of −.16.[13] Thus, our results imply substantially smaller effects than those reported in Levitt and instead are comparable to those of Marvell and Moody. These comparisons are misleading, however, as our departure from Levitt's results and our accordance with those in Marvell and Moody are driven largely by the difference in periods analyzed. While the results in Table 4 are based on a panel data set spanning 1978–2004,

[12] We calculate elasticities in the following manner. We divide the coefficient estimate by the average crime rate for the entire sample. We then divide this ratio by 1 divided by the average incarceration rate. Thus, the elasticity can be interpreted as the average scale-independent effect at the means of the sample.

[13] Marvell and Moody (1994) do not report separate elasticity estimates for overall violent and property crime. However, given the greater frequency of property crime, their overall effect is roughly consistent with the elasticity estimates that we find using our entire panel data set.

Table 4

Effect of Changes in Incarceration Rates on Changes in Overall Violent and Property Crime Rates: Full State-Level Panel

	ΔViolent Crime Rate								ΔProperty Crime Rate							
	Specification (1)				Specification (2)				Specification (1)				Specification (2)			
	OLS		IV		OLS		IV		OLS		IV		OLS		IV	
ΔIncarceration Rate	.048	(.081)	−.116	(.165)	.038	(.089)	−.230	(.176)	−.994	(.505)	−2.315	(.537)	−1.109	(.535)	−3.272	(.578)
ΔPopulation, by age (%)																
0–17	−5.704	(5.823)	−5.753	(6.041)	0.195	(6.357)	−0.036	(6.653)	41.477	(48.255)	41.084	(49.260)	89.019	(56.187)	87.156	(56.769)
18–24	−5.281	(7.092)	−4.649	(7.114)	2.450	(7.781)	3.272	(7.874)	7.121	(51.981)	12.184	(49.662)	65.427	(62.569)	72.079	(58.584)
25–44	−6.889	(9.350)	−6.593	(9.536)	−3.207	(10.038)	−2.702	(10.241)	99.559	(70.105)	101.929	(70.434)	128.378	(76.514)	132.459	(76.023)
45–64	−2.115	(7.721)	−1.028	(7.634)	−5.662	(8.290)	−3.820	(8.766)	105.093	(74.227)	113.797	(72.227)	82.560	(66.109)	97.458	(66.382)
ΔUnemployment Rate	−1.776	(1.398)	−2.028	(1.460)	−1.813	(1.446)	−2.183	(1.530)	27.294	(9.799)	25.269	(9.976)	27.850	(9.722)	24.855	(10.142)
ΔPoverty Rate	−.955	(.961)	−.938	(.946)	−.774	(.941)	−.735	(.915)	4.522	(4.413)	4.659	(4.584)	6.019	(4.203)	6.329	(4.574)
ΔBlack (%)	.339	(.322)	.494	(.306)	.382	(.300)	.643	(.271)	3.810	(1.676)	5.050	(1.499)	3.874	(1.384)	5.988	(1.360)
ΔPer Capita Income	−.001	(.003)	−.002	(.003)	.003	(.003)	.002	(.004)	−.065	(.024)	−.073	(.023)	−.024	(.029)	−.032	(.028)
State effects	No		No		Yes		Yes		No		No		Yes		Yes	
R^2	.471		.468		.487		.481		.503		.496		.532		.516	
Implied elasticity at the mean	.023		−.057		.019		−.113		−.064		−.151		−.072		−.213	

Note. Robust standard errors, clustered by state, are in parentheses. All models include a constant term and year fixed effects and are weighted by state-level population. N = 1,321. OLS = ordinary least squares; IV = instrumental variables.

Table 5

Effect of Changes in Incarceration Rates on Changes in Individual Crimes: Full State-Level Panel

	Specification (1)		Specification (2)	
Dependent Variable	OLS	IV	OLS	IV
ΔMurder	−.002 (.002)	−.005 (.006)	−.001 (.002)	−.005 (.006)
ΔRape	−.005 (.004)	−.029 (.011)	−.005 (.004)	−.033 (.015)
ΔRobbery	−.025 (.035)	−.165 (.111)	−.028 (.035)	−.227 (.136)
ΔAssault	.079 (.055)	.082 (.163)	.072 (.060)	.035 (.192)
ΔBurlgary	−.398 (.124)	−.857 (.273)	−.414 (.129)	−1.064 (.394)
ΔLarceny	−.498 (.286)	−1.182 (.353)	−.573 (.308)	−1.720 (.296)
ΔMotor vehicle theft	−.097 (.137)	−.275 (.253)	−.122 (.138)	−.487 (.274)
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes

Note. Robust standard errors, clustered by state, are in parentheses. All models include a constant term and control variables for changes in the age structure, the percentage of minority residents, the percentage of poor residents, the state unemployment rate, and state per capita income. All models include year fixed effects and are also weighted by state-level population. OLS = ordinary least squares; IV = instrumental variables.

Levitt's analysis is based on panel data spanning 1971–93, and Marvell and Moody analyze the period 1971–89. As we show below, we find considerably larger effects when we restrict our sample to an earlier time period. Moreover, unlike Marvell and Moody, we consistently find strong evidence that the crime-prison elasticities estimated by OLS are severely biased toward zero.

Table 5 presents comparable estimates for each of the seven individual felony offenses listed in Table 3. Again, we report results from four separate models. Here, however, we report only the coefficients on the change in incarceration rates. With the exception of the assault rate models, instrumenting the change in incarceration rates with our predicted change in incarceration rates yields more negative effects of incarceration rates on crime rates, with most point estimates substantially larger and statistically distinguishable from the OLS results. None of the coefficients in the murder rate models are statistically significant, although all are negative. The IV results for rape suggest much larger (roughly six times larger) effects of changes in incarceration rates on the rape rate than those using the OLS specification. We find similar results for robbery, burglary, larceny, and auto theft, but no measurable effect for assault. Finally, incorporating state fixed effects in the model specification increases the magnitude of the point estimates in nearly all cases.

The results in Table 5 can be used to compare our estimation results to those from the literature on pure incapacitation effects that attempts to gauge crimes avoided through retrospective inmate surveys. Recall that although the estimates in this literature range from 10 crimes prevented to more than 100 crimes per additional inmate, the most careful assessments of this research suggest a range of estimates between 10 and 20 incidents per year of prison served. Since both the dependent and independent variables used in the models in Tables 4 and 5 are expressed per 100,000 state residents, the coefficients on the change in the incarceration rate can be interpreted as the average effect of putting one more person in prison for a year. Thus, summing the coefficients for the seven crime categories in Table 5 provides an estimate of the number of part I felony offenses prevented by putting one more person in prison.

One problem with this estimate concerns the fact that the UCR data are based on crimes reported to the police, and with the exception of murder, reporting rates are considerably lower than 1 for all crimes. However, with crime-specific data on reporting rates, one can easily inflate the point estimates by dividing by the proportion of incidents reported to the police.

Using the IV estimates in specification (2) in Table 5, and accounting for the underreporting of most crimes,[14] we find that, on average, a one-person increase in the prison population prevents .005 murder, .1 rape, .04 robbery, 0 assaults, 2.1 burglaries, 6.3 larcenies, and .6 motor vehicle theft. In total, this constitutes 9.4 fewer part I felony offenses for each additional inmate. Thus, our model estimates using the entire sample are on the low end of the range of estimates from the pure-incapacitation literature.[15]

Before turning to our subperiod analysis, we assess the sensitivity of our main results to some of our key estimation choices. First, we employ an instrumental variable that has been smoothed in an attempt to identify permanent changes in the parameters determining the steady-state incarceration rate. Second, all regressions in our tables are weighted by state-level population. Table A1 presents regression results for violent and property crime models as well as for each of the individual part I offenses. Comparing the population-weighted results with the smoothed and unsmoothed instruments shows that our main results are not sensitive to this choice. Smoothing the instrument generally yields smaller standard errors. However, very similar patterns in terms of coefficient magnitude and statistical significance are observed for both sets of results.

The results are sensitive to weighting by state population. In particular, the unweighted standard errors are considerably larger than the standard errors even with population weights. Consequently, the greater imprecision yields many fewer significant effects. In defense of the weighting, we believe that larger states should receive greater weight for several reasons. First, crime and incarceration rates are measured more precisely, as are the underlying transition probabilities, and hence the weighted models are more efficient. Because we report standard errors that are robust to heteroskedasticity as well as within-state serial correlation, our inference is valid even if the simple weighting scheme misspecifies the nature of

[14] Rennison (2001) presents, for 1993–2000, estimates from the National Criminal Victimization Survey on the proportion of offenses reported to the policy by crime victims. Averaging her eight annual estimates yields average reporting rates of .325 for rape, .572 for robbery, .553 for aggravated assault, .502 for burglary, .262 for larceny, and .788 for motor vehicle theft. In the aforementioned calculations, we use these reporting rates to inflate the marginal effects of incarceration. We assume that the police are aware of all murders.

[15] Note that the recent incapacitation research by Owens (2009) suggests even smaller effects (on the order of 2–3 crimes per year).

Table 6

First-Stage Effect of the Predicted Change in Incarceration Rates Based on Last-Period Shock on the Current Change in Incarceration Rates, by Subperiod

	(1)	(2)	(3)	(4)
1978–90:				
Predicted ΔIncarceration	.545 (.074)	.513 (.076)	.424 (.088)	.182 (.105)
F-statistic [*p*-value]	54.28 [<.0001]	45.78 [<.0001]	23.43 [<.0001]	2.98 [.090]
1991–2004:				
Predicted ΔIncarceration	.625 (.142)	.594 (.164)	.564 (.166)	.476 (.203)
F-statistic [*p* -value]	19.21 [<.0001]	13.06 [<.0007]	11.51 [<.0014]	5.52 [<.023]
Control variables	No	Yes	Yes	Yes
Year effects	No	No	Yes	Yes
State effects	No	No	No	Yes

Note. The dependent variable is ΔIncarceration Rate. Robust standard errors, clustered by state, are in parentheses. All models are weighted by state-level population. *F*-statistics are from a simple test of the significance of the instrument in the first-stage regression.

the heteroskedasticity in the data (Cameron and Trivedi 2005). Second, weighting by state population yields parameter estimates that better reflect the typical relationship between state-level crime and incarceration rates experienced by the average U.S. resident. Finally, over the past decade or so, nearly all of the panel data research on the determinants of crime weight the regression by state population. Hence, we believe that our preferred estimates in Tables 4 and 5 are more in line with current empirical practice.

6. Estimation Results by Subperiod

The estimates in the Section 5 constrain the average effect of incarceration rates on crime rates to be constant across all time periods and states. Given the substantial increase in U.S. incarceration rates since the mid-1970s, this specification choice is likely to be too restrictive. Assuming that the most criminally active are incarcerated first, either through the deliberate targeting of resources by the criminal justice system or through the most active being caught first, we would expect that the marginal crime-abating effect of an additional inmate would be lower in later years (when the incarceration rate is higher) than in earlier years.

Here we assess whether the crime-prison effects vary by period. To do so, we stratify our panel into two subpanels covering the periods 1978–90 and 1991–2004 and estimate separate models for each period. We begin with an analysis of the strength of the first-stage relationship between our predicted change in incarceration rates and the actual change for these subperiods. Table 6 presents the results from several first-stage models, with regression specifications corresponding to those employed in Table 2. We report only the coefficient on the predicted change in incarceration rates. Specifications (3) and (4) correspond to the first-stage models that we subsequently use in the crime-prison models.

Table 7
Effect of Changes in Incarceration Rates on Changes in Violent and Property Crime Rates, by Subperiod

	Specification (1)		Specification (2)	
Dependent Variable	OLS	IV	OLS	IV
ΔViolent Crime Rate:				
Marginal effect:				
1978–90	.135 (.172)	−.529 (.712)	.028 (.215)	−2.534 (1.867)
1991–2004	.022 (.086)	−.076 (.167)	−.024 (.084)	−.321 (.133)
Implied elasticity:				
1978–90	.042	−.166	.009	−.794
1991–2004	.014	−.048	−.015	−.206
ΔProperty Crime Rate:				
Marginal effect:				
1978–90	−2.192 (1.004)	−4.163 (3.240)	−2.422 (.925)	−11.414 (6.105)
1991–2004	−.666 (.528)	−1.832 (.632)	−.799 (.485)	−2.693 (.731)
Implied elasticity:				
1978–90	−.083	−.157	−.092	−.432
1991–2004	−.062	−.170	−.074	−.250
State effects	No	Yes	No	Yes

Note. Robust standard errors, clustered by state, are in parentheses. All models include a constant term and control variables for changes in the age structure, the percentage of minority residents, the percentage of poor residents, the state unemployment rate, and state per capita income. All models include year fixed effects are also weighted by state-level population. OLS = ordinary least squares; IV = instrumental variables.

In all four specifications and for both time periods, the predicted change in incarceration rate has a positive and significant effect on the actual change in incarceration rate, although the relationship is weak for specification (4) in the early time period. We find a better first-stage fit for the later time period, which suggests greater temporary variation in the underlying determinants of incarceration in the early period. Moreover, the degree to which the prediction overestimates the actual change is greater in the earlier period.

Table 7 presents OLS and IV estimates of the effect of changes in incarceration on changes in violent and property crime rates for each period. The specifications correspond exactly to those used in Table 4, although here we report only the coefficients on incarceration. Table 7 also reports the implied elasticities at the sample means. In both periods the IV estimates for violent crime are more negative than the OLS estimates (which are all near zero and statistically insignificant), although the IV point estimates are poorly measured and statistically insignificant in all but specification (2) for the later period. The IV estimates for the earlier period are considerably larger than those for the later period. When year effects for the 1978–90 model are included, the results suggest that each additional prison-year served prevents roughly .5 of a violent crime, while including state effects yields the much larger estimate of 2.5 violent crimes avoided for each year served. These point estimates correspond to violent crime–prison elasticities of −.166 and −.794 (consistent with the findings from Levitt

Table 8

Effect of Changes in Incarceration Rates on Changes in Individual Part I Felony Offenses, by Subperiod

	Specification (1)		Specification (2)	
Dependent Variable	OLS	IV	OLS	IV
Murder:				
1978–90	.006 (.004)	−.002 (.032)	.002 (.004)	−.038 (.072)
1991–2004	−.003 (.002)	−.006 (.003)	−.003 (.002)	−.006 (.004)
Rape:				
1978–90	−.011 (.009)	−.075 (.027)	−.005 (.011)	−.202 (.112)
1991–2004	−.002 (.005)	−.019 (.008)	−.002 (.005)	−.021 (.014)
Robbery:				
1978–90	−.141 (.090)	−.793 (.461)	−.216 (.109)	−2.555 (1.356)
1991–2004	−.010 (.037)	−.086 (.089)	−.043 (.028)	−.257 (.148)
Assault:				
1978–90	.281 (.120)	.341 (.388)	.248 (.138)	.262 (1.121)
1991–2004	.037 (.050)	.036 (.137)	.023 (.055)	−.037 (.143)
Burglary:				
1978–90	−1.031 (.424)	−2.731 (1.375)	−1.078 (.385)	−6.769 (3.415)
1991–2004	−.224 (.126)	−.474 (.169)	−.232 (.105)	−.514 (.198)
Larceny:				
1978–90	−1.076 (.446)	−1.249 (1.516)	−1.081 (.396)	−2.627 (3.519)
1991–2004	−.328 (.289)	−1.027 (.356)	−.445 (.293)	−1.674 (.365)
Motor vehicle theft:				
1978–90	−.085 (.289)	−.183 (1.039)	−.263 (.335)	−2.018 (2.249)
1991–2004	−.114 (.133)	−.331 (.199)	−.123 (.109)	−.505 (.338)
State effects	No	Yes	No	Yes

Note. Robust standard errors, clustered by state, are in parentheses. All models include a constant term and control variables for changes in the age structure, the percentage of minority residents, the percentage of poor residents, the state unemployment rate, and state per capita income. All models include year fixed effects and are also weighted by state-level population. OLS = ordinary least squares; IV = instrumental variables.

[1996]), although the large standard errors suggest a much wider confidence interval.

In contrast, the violent crime results for the later period are considerably more modest and significant only when state effects are included in the specification. In model 2, the IV estimates suggest that each prison-year served prevents .3 violent crime, which corresponds to an elasticity estimate of −.2.

We observe similar yet more precise patterns for property crime. Again, IV estimates are generally much more negative than the OLS estimates. For the earlier period, the IV-estimated effects suggest that each prison-year served prevents between 4 and 11 property crimes, which corresponds to elasticity estimates of −.157 and −.432. Again, this range includes the elasticity estimates in Levitt (1996) (−.26 to −.32). The point estimates are considerably smaller for the later period.

Table 8 presents corresponding crime-specific results. In nearly all cases, IV estimates are more negative than OLS estimates, and the estimated effect sizes are larger for the earlier period than the later period. To summarize these results, we again estimate the total number of crimes avoided by an additional prison-year served, adjusted for crime-specific underreporting rates in the UCR data. To do so, we make use of the estimation results in the final column of Table 8, which uses the most liberal specification of the IV model. The results indicate that for the period 1978–90, each additional inmate prevented approximately 30 part I felony offenses. The comparable value for the period 1991–2004 is 8.3. Thus, the marginal crime-fighting effect of an additional inmate has indeed declined substantially in recent years.

7. Conclusion

This study has several important findings and contributions. First, to isolate exogenous variation in incarceration, we use the dynamic adjustment path of aggregate incarceration rates induced by shocks to underlying determinants of incarceration rates. Our theoretical prediction regarding subsequent 1-year changes in incarceration rates provides a strong instrument for actual changes in incarceration rates. Moreover, the strategy permits us to estimate separate effects for the United States, stratified by period, because the instrument does not depend on variation in specific states or periods.

Second, we find that breaking the simultaneity between incarceration and crime rates yields substantially larger estimates of the effects of incarceration on crime than those estimated in simple OLS models. When restricted to earlier periods, our corrected estimates are in accord with those of Levitt (1996), who analyzes a similar period yet uses an entirely different identification strategy.

Third, we find that the effect on crime rates of incarcerating one more inmate has declined drastically over the past quarter century. When we split the sample into two equal periods, we find crime-prison effects for the later period that are less than one-third the size of those for the earlier period. For 1978–90, we estimate that each additional prison-year served prevented approximately 30 index crimes. For the period 1991–2004, the comparable value is eight. Moreover, this decline in level effects corresponds to substantial declines in crime-prison elasticities, which suggests that the constant-elasticity specification often used in previous research underestimates the degree to which the crime-abating effects of incarceration decrease with scale.

This large decline in the marginal effect of an inmate suggests that the most recent increases in incarceration have been driven by the institutionalization of many inmates who, relative to previous periods, pose less of a threat to society. Indeed, given the much lower crime-abating effects for the most recent period, it is likely the case that for many recent inmates, the benefits to society in terms of crime reduction are unlikely to outweigh the explicit monetary costs of housing and maintaining an additional inmate. Moreover, once we accounts for the additional external costs of incarceration, such as the adverse effects on the families of inmates, the effects on victimizations behind bars, the effects on additional HIV/AIDS infections (Johnson and Raphael 2010), and the potential effects on the long-term employment prospects of former inmates, the benefit-cost ratio on the margin is likely to be substantially less than one.

Appendix

Incorporating Behavioral Responses into the Model of Incarceration and Crime

The model presented in Section 3 assumes that the underlying transition parameters in our aggregate model of crime and incarceration do not respond to one another, either instantaneously or with a dynamic lag. Here we sequentially relax some of the behavioral assumptions that we made and assess how this impacts the interpretation of our results.

We first maintain the assumption that underlying criminality is unresponsive to either the threat of incarceration (captured by p) or the severity of punishment (captured by the release parameter θ). We begin by considering the case in which the likelihood of being apprehended and incarcerated varies with the degree of underlying criminality—that is, where $p = p(c)$. One might hypothesize a priori that p may be either increasing or decreasing in the degree of criminality. If policy makers respond to an increase in criminal behavior by greatly increasing resources devoted to policing and corrections, we would expect that increases in criminality would lead to increases in the risk of incarceration. On the other hand, to the extent that increases in criminality dilute enforcement resources, dp/dc may be negative.

With regard to our identification strategy, in isolation such a policy reaction should not compromise the exogeneity of our proposed instrument, although the timing of this response may impact the strength of our first-stage relationship. To see this, first consider the case in which p responds instantaneously to changes in c. For $dp/dc > 0$, an increase in criminality will cause an instantaneous increase in the incarceration risk. The higher value of p will translate into a lower initial spike in crime, a larger initial increase in incarceration, and a speedier adjustment of incarceration and crime rates to their new equilibrium levels. The opposite would apply when p decreases in c. In either case, the change in incarceration rate subsequent to the initial change still provides exogenous variation in incarceration and thus the IV strategy outlined above still identifies a causal incapacitation effect.

However, the incarceration risk parameter is unlikely to respond instantaneously, since policy makers are likely to learn of an increase in c only with time and since the budgetary process impedes instantaneous reaction to new problems. Such a delayed response will impact the first-stage relationship between the predicted change in incarceration in equation (10) and actual changes. To see this, suppose that at $t = 0$ the risk of incarceration is p_0 but that it increases to p_1 at the beginning of period 1 (fully 1 period after the increase in criminality modeled above). Given that the incarceration risk remained constant for the first period following the shock, the initial change in the incarceration rate remains equal to that described in the first line of equation (10). However, the new increase in incarceration between periods 1 and 2 becomes

$$\Delta S_{2,1} = \Delta + (S^{\star}_{2,2} - S^{\star}_{2,1})(c_1 p_0 + \theta)S_{2,1|\text{constant } p} + c_1(p_1 - p_0)(S^{\star}_{2,2} - S_{2,1}), \tag{A1}$$

where the first term is the predicted change in incarceration holding the incarceration risk constant (the second line in equation [10] and our proposed instrument), $S^{\star}_{2,2} = c_1 p_1/(c_1 p_1 + \theta)$ is the new long-run equilibrium incarceration rate with the elevated incarceration risk, $S^{\star}_{2,1} = c_1 p_0/(c_1 p_0 + \theta)$ is the equilibrium incarceration rate given last period's parameters, and $S_{2,1}$ is the actual incarceration rate in period 1. The second term in equation (A1) adjusts our previous estimate of the increase in incarceration for this period for the change in the long-term equilibrium rate, while the third terms reflects the instantaneous effect of the increase in the incarceration risk parameter p. When p is increasing in c, both terms are positive. Thus, our proposed instrument will systematically underestimate the actual change in incarceration rate. If p is a decreasing function of c, then the opposite would hold. Regardless, our proposed instrument is still a component of the change in incarceration rate, and thus a first-stage relationship should exist. Moreover, because the criminality parameter is assumed to be constant across periods, this particular behavioral response will not induce correlation between the instrument and the second-stage error term.

The impact of a change in sentence length on our instrument requires a dynamics analysis, since a change in sentence length today will not impact changes in incarceration rates until today's cohort of admitted inmates reaches their counterfactual release dates under the prior sentencing regime. In practice, inmates sentenced to state or federal prison receive sentences of at least 1 year, although those being sent back for parole violation often serve terms that may fall short of a full year (Raphael and Weiman 2008). Here we work through the effect of an increase in sentence length in response to an increase in criminality, operating under the assumption that an increase in c at time $t = 0$ does not impact release rates until $t = 1$.

If policy makers increase sentence severity in response to an increase in criminality, this implies that period release rates decrease in the once-lagged value of c, or $\theta'_t(c_{t-1}) < 0$. Again, since θ does not change instantaneously in response to a change in the criminality parameter, the period 1 change in incarceration rates will not differ from that in the first line of equation (10). The change in incarceration between periods 1 and 2, however, will differ by an identifiable quantity. To be specific, assuming that the release probability increases from θ_0 at $t = 0$ to θ_1 at $t = 1$, the implied change in the incarceration rate between periods 1 and 2 will be

$$\Delta S_{2,1} = \Delta S_{2,1|\text{constant }\theta} + (\theta_0 - \theta_1)S_{2,1}. \tag{A2}$$

The first term on the right-hand side of equation (A2) is our previous estimate of the period change in incarceration rates, assuming no responsiveness of θ to changes in c, while the second term reflects an additional component associated with the lower release rate at $t = 1$ (which will be positive for $\theta_1 < \theta_0$). Thus, allowing sentence severity to respond punitively to increases in criminality suggests that our proposed instrument will underestimate the increase in incarceration. Nonetheless, the instrument again predicts a component of the change and thus should be able to identify significant variation in the future change in incarceration rates. Moreover, we are holding the criminality parameter constant by assumption, and thus the instrument will not be correlated with the error term in the second-stage equation.

Allowing criminality and corrections policy to respond reciprocally complicates the analysis somewhat and does indeed bias the second-stage estimate of our structural estimate of the effect of incarceration on crime, if we interpret this estimate narrowly as a pure incapacitation effect. However, if we broaden the interpretation of the effect that we are estimating to allow for a general deterrence effect in addition to an incapacitation effect, our instrument is still valid. To see this, we begin with a simple model whereby criminality and enforcement adjust instantaneously to one another. We then think through a likely dynamic structure of this response and how it will impact our estimation strategy.

Suppose that criminality is determined by two factors: a variable x measuring criminogenic influences that are unresponsive to policy and the overall incarceration risk p. Assume further that criminality is determined according to the additively separable function $c = h(x) + f(p)$, where $h'(x) > 0$ and $f'(p) < 0$. We also assume that the incarceration risk is a monotonically increasing function of the criminality parameter $h(x) + f(p) = g^{-1}(p)$—that is, $p = g(c)$, where $g'(c) > 0$. The assumption that p is increasing in c best describes the recent history of corrections policy in the United States.[16] The equilibrium incarceration risk is defined by the condition

$$h(x) + f(p) = g^{-1}(p), \tag{A3}$$

where $g^{-1}(p)$ is the inverse of the function $g(c)$. Totally differentiating condition (A3) with respect to x and the equation for c gives the response of p and c in response to a change in x:

[16] The period with perhaps the strongest evidence in favor of an overall increase in criminality occurred during the later 1980s and early 1990s (the period corresponding to the crack cocaine epidemic). During that period, prison admission rates increased steeply relative to previous levels, and sentences were enhanced at the federal level and in many states for nonviolent drug crimes and for other felony offenses. For detailed discussions of sentencing policy in the United States, see Tonry (1996) and Jacobson (2006).

$$\frac{dp}{dx} = \frac{h'(x)}{g^{-1\prime}(x) - f'(p)},$$
$$\frac{dc}{dx} = h'(x)\left[1 + \frac{f'(p)}{g^{-1\prime}(p) - f'(p)}\right]. \tag{A4}$$

For $g'(c) > 0$, both derivatives in equation (A4) are positive. Thus, a positive shock to underlying crime fundamentals leads to an equilibrium increase in both criminality and the incarceration risk.

The implications of this bidirectional responsiveness for our identification strategy will again depend on whether these adjustments occur instantaneously or over time. If a shock to c causes an instantaneous adjustment of the incarceration risk and criminality to new equilibrium levels, then our identification strategy is not affected. Future increases in incarceration will be based on stable behavioral and policy parameters, and our proposed instrument will identify exogenous variation in future changes in incarceration rates.

However, the reaction of policy to a change in criminality is likely to be slower than the reaction of criminal behavior to changes in enforcement, since changes in policy will be slowed to some degree by the timing of the budgetary process. Moreover, if policy makers choose an optimal enforcement level based on their best estimate of current criminality, equilibrium values of p and c may not be reached after only a multiperiod adjustment process. Here we consider the implications for our identification strategy of the specialized case where c responds instantaneously to changes in p but where p responds to changes in c with a 1-period lag. Since our chosen instrument will be used to predict changes in incarceration between periods $t = 1$ and $t = 2$, we consider only the effect of this reciprocal responsiveness for the first two changes in crime and incarceration rates following a shock to c.

Assume that at $t = 0$, criminality increases from c_0 to c_1. Given the lagged responsiveness of the incarceration risk, p will not increase until period 1. At that time, the parameter increases from p_0 to p_1. Since criminality responds instantaneously, the criminality parameter will decrease from c_1 to c_2, as suggested by the simple model in equations (A3) and (A4).[17] For this example, the first two changes in observed crime rates are given by the equations

$$\Delta C_0 = -c_1 \Delta S_{2,0} + (c_1 - c_0)(1 - S^{*}_{2,0})$$
$$\Delta C_1 = -c_2 \Delta S_{2,1} + (c_2 - c_1)(1 - S_{2,1}). \tag{A5}$$

Note that the change in the crime rate between periods 0 and 1 is equivalent to the change reported in equation (11), where we assumed away behavioral

[17] To be sure, c_2 will not be an equilibrium value for criminality but rather the optimal behavioral response to the incarceration risk associated with p_1. It will take several additional policy iterations and reciprocal behavioral responses to reach steady-state values for the parameters. However, since we are primarily interested in instrumenting the change in incarceration between periods 1 and 2, we do not need to consider subsequent changes.

responses of the parameters. The change in crime rates between periods 1 and 2, however, has changed, with an alternative coefficient on the change in incarceration ($-c_2$ instead of $-c_1$) and an additional term capturing the effect on crime of the decrease in criminality from c_1 to c_2. This second term is basically the general deterrence effect of an increase in the incarceration risk.

Since we cannot observe the actual values for the criminality parameter, this general deterrence component will be swept into the error term in the second-stage crime equation (given by the second line in equation [A5]). Since the decreases in criminality in period 1 are driven by the response of policy to the increases in criminality in period 0, the change in criminality between periods 0 and 1 will be negatively correlated with the change in criminality between periods 1 and 2. Our proposed instrument is increasing in the initial increase in criminality. Thus, our instrument will be negatively correlated with the deterrence effect contained in the error term of the second-stage regression, which causes a negative bias to our estimate of the incapacitation effect of prison (that is, our elasticity estimates will be too large).

Despite this change, however, we are still able to interpret the structural estimate from the second stage as a causal effect of corrections policy on crime, although we must change our interpretation somewhat. With a lagged reciprocal response, the second-stage estimate of the coefficient on the change in incarceration will reflect both the incapacitation effect and the general deterrence effect of prison on crime (through the correlation between the predicted incarceration change and the general deterrence component in the error term). Thus, although we are unable to disentangle the separate avenues by which prison is likely to impact criminality activity, our proposed IV estimate does permit estimation of a cumulative impact.

With regard to our first-stage prediction, the predicted increase in incarceration with stable parameters may either overestimate or underestimate the actual increase. With the stated sequence of criminality and incarceration risk values listed above, the actual increase in incarceration between periods 1 and 2 is given by the equation

$$\Delta S_{2,1} = \Delta S_{2,1|\text{constant } p, c} + (S_2^* - S_1^*)(c_1 p_0 + \theta) + (S_2^* - S_1)(c_2 p_1 - c_1 p_0), \quad \text{(A6)}$$

where the first term is the predicted change in incarceration between periods 1 and 2, assuming stable parameters following the initial crime shock, where $S_2^* = c_2 p_1/(c_2 p_1 + \theta)$ is the new equilibrium incarceration rate after the period 1 response of criminality to the change in the incarceration risk, $S_1^* = c_1 p_0/(c_1 p_0 + \theta)$ is the equilibrium rate after the initial criminality shock but preceding the incarceration risk response, and S_1 is the actual incarceration rate at the beginning of period 1. The second and third terms in this model will both be either positive or negative depending on the relative size of $c_2 p_1$ and $c_1 p_0$. If $c_2 p_1 > c_1 p_0$, the increase in the incarceration risk is so large relative to the subsequent decrease in criminality that the equilibrium incarceration rate increases. In this instance, the instrument will underestimate the actual increase in incarceration. On the other hand, if $c_2 p_1 < c_1 p_0$, then the decline in criminality is sufficiently large relative to the increase in the incarceration risk that the equilibrium rate declines and both of the second terms in equation (A6) are negative. In this instance, our nonbehavioral prediction will overestimate the actual change in the incarceration rate.

There is one potential behavioral response on the part of the potentially criminal that may bias our results, even when interpreted to represent a total effect of incarceration on crime. Suppose that current increases in criminal behavior cause subsequent decreases in criminality because of a revulsion on the part of the general public to the consequences of the crime spike. Such responses suggest that changes in the criminality parameter may exhibit negative serial correlation. An example of such a revulsion effect may be a drug epidemic that runs it course when a younger generation is reluctant to use a drug that devastated the lives of the preceding generation. This type of serial correlation would create a negative correlation between our instrument and the corresponding change in the crime rate, for reasons similar to those given in the discussion of equation (A6). Here however, the decline in the crime rate cannot be attributed to a general deterrence effect of prison, since the decline would occur regardless of a change in the chance of punishment. In this case, there would be a spurious reduced-form relationship between the instrument and the change in the crime rate.

Although such changes in behavior are certainly possible, that spikes in criminal behavior driven by the introduction of a new drug or other variants of crime shocks would correct themselves within a year or two is unlikely and contrary to recent history. Nonetheless, we acknowledge this threat to the validity of our estimation strategy.

Table A1

Comparison of Point Estimates of Crime-Prison Effects

	OLS		IV			
			Smoothed Instrument		Unsmoothed Instrument	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
All violent crime:	.038 (.089)	.244 (.131)	−.230 (.175)	.040 (.693)	−.203 (.195)	−.128 (.465)
Murder	−.001 (.002)	.006 (.003)	−.005 (.006)	.049 (.044)	−.009 (.006)	.018 (.023)
Rape	−.005 (.004)	.004 (.004)	−.033 (.014)	−.118 (.076)	−.030 (.015)	−.047 (.063)
Robbery	−.028 (.035)	.037 (.040)	−.227 (.136)	−.155 (.354)	−.247 (.154)	−.132 (.274)
Assault	.072 (.060)	.197 (.091)	.035 (.192)	.264 (.324)	.083 (.176)	.033 (.187)
All property crime:	−1.109 (.535)	−.048 (.479)	−3.272 (.577)	−2.296 (2.759)	−3.245 (1.058)	−2.765 (2.512)
Burglary	−.414 (.128)	−.140 (.138)	−1.064 (.394)	−.641 (.983)	−1.278 (.578)	−.992 (.879)
Larceny	−.573 (.308)	−.040 (.224)	−1.720 (.296)	−.997 (1.417)	−1.556 (.419)	−1.488 (1.267)
Motor vehicle theft	−.122 (.138)	.132 (.126)	−.487 (.274)	−.658 (.530)	−.411 (.347)	−.284 (.484)

Note. Robust standard errors, clustered by state, are shown in parentheses. Values are coefficients from regressions of the first difference of each crime rate on the first difference of the incarceration rate. Each specification includes complete sets of state and year fixed effects as well as control variables for changes in the age structure, the percentage of minority residents, the percentage of poor residents, the state unemployment rate, and state per capita income. The weighted regressions weight each observation by state-level population. OLS = ordinary least squares; IV = instrumental variables.

References

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* New York: Cambridge University Press.

DiIulio, John J., and Anne Piehl. 1991. Does Prison Pay? The Stormy National Debate over the Cost-Effectiveness of Imprisonment. *Brookings Review* 9(4):28–35.

Donohue, John J., III, and Peter Siegelman. 1998. Allocating Resources among Prisons and Social Programs in the Battle against Crime. *Journal of Legal Studies* 27:1–43.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova. 2009. The Deterrent Effects of Prison: Evidence from a Natural Experiment. *Journal of Political Economy* 117:257–80.

Jacobson, Michael. 2005. *Downsizing Prisons: How to Reduce Crime and End Mass Incarceration.* New York: New York University Press.

Johnson, Rucker, and Steven Raphael. 2010. The Effect of Male Incarceration Dynamics on AIDS Infection Rates among African-American Women and Men. *Journal of Law and Economics* 52:251–93.

Kessler, Daniel, and Steven D. Levitt. 1999. Using Sentence Enhancements to Distinguish between Deterrence and Incapacitation. *Journal of Law and Economics* 42:343–63.

Lee, David S., and Justin McCrary. 2005. Crime, Punishment, and Myopia. Working Paper No. 11491. National Bureau of Economic Research, Cambridge, Mass.

Levitt, Steven D. 1996. The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Legislation. *Quarterly Journal of Economics* 111:319–51.

———. 1998. Juvenile Crime and Punishment. *Journal of Political Economy* 106:1156–85.

Marvell, Thomas, and Carlysle Moody. 1994. Prison Population Growth and Crime Reduction. *Journal of Quantiative Criminology* 10:109–40.

Owens, Emily G. 2009. More Time, Less Crime? Estimating the Incapacitative Effects of Sentence Enhancements. *Journal of Law and Economics* 52:551–79.

Raphael, Steven, and David Weiman. 2008. The Impact of Local Labor Market Conditions on the Likelihood That Parolees Are Returned to Custody. Pp. 304–22 in *Barriers to Reentry? The Labor Market for Released Prisoners in Post-industrial America,* edited by Shawn Bushway, Michael A. Stoll, and David F. Weiman. New York: Russell Sage Foundation.

Rennison, Callie M. 2001. *Criminal Victimization 2000: Changes 1999–2000 with Trends 1993–2000.* NCJ 187007. Washington, D.C.: Bureau of Justice Statistics.

Spelman, William. 1994. *Criminal Incapacitation.* New York: Plenum Press.

———. 2000. What Recent Studies Do (and Don't) Tell Us about Imprisonment and Crime. Pp. 419–94 in *Crime and Justice: A Review of the Research,* edited by Michael Tonry. Chicago: University of Chicago Press.

Tonry, Michael. 1996. *Sentencing Matters.* Oxford: Oxford University Press.

Webster, Cheryl M., Anthony N. Doob, and Franklin E. Zimring. 2006. Proposition 8 and Crime Rates in California: The Case of the Disappearing Deterrent. *Criminology and Public Policy* 5:417–48.

Western, Bruce. 2006. *Punishment and Inequality in America.* New York: Russell Sage Foundation.