Dynamic Perspectives on Crime*

Justin McCrary
UC Berkeley, NBER

This paper was prepared for the *Handbook of the Economics of Crime*.

September 25, 2009

*I think David Lee for collaboration on related projects and Pat Kline and Demian Pouzo for unusually helpful comments. Mike Elsby provided a set of extremely detailed and insightful comments on an earlier draft. I am responsible for any errors.

1 Introduction

Economists seeking to understand how crime might respond to policy interventions typically turn to the model of crime due to Becker (1968). In this model, crime is viewed as a point-in-time bet, and crime occurs when the expected utility of taking the bet is greater than the expected utility of turning it down. Another workhorse of the economic analysis of crime is the static time allocation model of Gronau (1977). This model has been adapted by several authors, including Grogger (1998), Lemieux, Fortin, and Frechette (1994), and Williams and Sickles (2002). The models of Becker and Gronau are static, in the sense that there is no explicit reference to the future consequences of apprehension.

Static models of criminal labor supply lead to important insights in many contexts, but can be awkward in contexts where dynamics are important. As one example, crimes with the greatest social costs are serious property and violent crimes. In every modern society, these crimes are punished by lengthy prison sentences rather than fines or instantaneous physical punishment. Thus, the disutility associated with apprehension for the most important crimes is experienced many periods after the utility gain associated with commitment of the crime. A second example of the importance of dynamic considerations pertains to intertemporal substitution of criminal activity. Over the last decade or so, a consensus has emerged in criminology that "hot spots policing"—i.e., a massive increase in police presence within

a geographically small area of a city where criminal activity has recently been high—is a highly effective strategy for crime reduction. However, to the extent that criminals are able to relocate activity to other time periods or other places, such strategies may be more effective at reducing crimes in one location temporarily than in reducing crimes in the aggregate over the medium- to long-term.

A static perspective is also limiting when it comes to the government's problem of controlling crime. Consider a state legislator, who is deciding whether to vote for a sentence enhancement bill, or whether to vote for a bill to subsidize hiring of police officers on the part of localities. Assuming modest magnitudes for deterrence elasticities, a vote for the sentence enhancement bill entails small current costs and large future costs, whereas a vote for the policing bill entails large current costs and medium future costs.

While these dynamic features of crime are interesting, they do not occupy a central place within the current economics of crime literature. My assessment is that this is an important gap. In this chapter, I review the literature on dynamics and crime. In light of the small size of this literature, more attention than is usual in a review article is devoted to potentially fruitful directions for future research.

The remainder of this chapter is organized as follows. The next section provides a brief overview of the existing small literature on dynamics and crime. I then review the simple dynamic model of crime from Lee and McCrary (2009). This model combines Becker's crime model with a job search

model. The penultimate section discusses the government's problem of minimizing the present discounted value of crime using adjustments to policing levels and sentence lengths, currently and in the future. This problem has many interesting dynamic features. I am not aware of any work on this important topic. The final section concludes.

2 Literature Review

The first dynamic treatment of crime of which I am aware is Flinn (1986).[1] Flinn models the proportion of time allocated to work and crime in each period. He assumes no borrowing and no savings and lets the probability of apprehension in period t be increasing in the amount of time devoted to crime in period t. Upon apprehension, the individual is incarcerated for a deterministic sentence length that can depend on prior criminal history. Wages in the legitimate market are either fixed, or increase with experience. Flinn shows that the model is capable of matching the age profile of crime, but does not seek to estimate or calibrate the model.

Lee and McCrary (2005, 2009) emphasize one of the most basic insights of a dynamic perspective on deterrence: if offenders have short time horizons, then it is hard to imagine punishment acting as an important deterrent. Empirically, most offenses occur at a time when the offender is experiencing diminished capacity. Nationally, at the time of arrest, 65 percent of arrestees

[1] Interestingly, this paper is part of an edited volume that is extremely prominent among criminologists, but somewhat obscure among economists.

have positive urinalysis tests for one of 5 major drugs: marijuana, cocaine, opiates, methamphetamines, or phencyclidine (PCP). Fully 21 percent test positive for the use of more than one such drug at the time of arrest.[2] Considering the possibility of both drug and alcohol use, it seems likely that a large fraction of offenders are prone to impulsive behavior. The mental state of the marginal offender at any given point in time thus may well be importantly different from that of a person contemplating decisions typically modeled in other areas of economics (how many years of schooling to obtain, lifecycle labor supply in the legitimate labor market, marriage and fertility, and so on), and indeed may be different than the mental state of the offender himself at other times (Strotz 1955).[3]

The time horizon of the offender is relevant because in every developed country around the world, serious crimes are punished by long prison sentences, measured in years or even decades. If the marginal offender has a short time horizon, it may be difficult to reduce his criminal propensity by threatening additional punishment. Importantly, however, even if the marginal offender has an extremely short time horizon, it may still be possible to reduce his criminal propensity using enhancements to the probability of apprehension (see Section 3 for details).

[2]Estimates from the Arrestee Drug Abuse Monitoring (ADAM) program of the National Institutes of Justice. The ADAM program uses probability sampling at 35 different sites scattered throughout the U.S. Numbers reported in text reflect arrestee-weighted averages of site-specific estimates and reflect author's calculations.

[3]For an explicit dynamic model of offense behavior under hyperbolic discounting, see Lee and McCrary (2005).

In their empirical work, Lee and McCrary (2009) use data from Florida to measure the rate of criminal involvement local to 18, when offenders are handled by the adult criminal justice system instead of the more lenient juvenile system. The estimates suggest only a 2 percent decline in the probability of offense upon transitioning to 18, when the expected period of detention, conditional on arrest, increases by roughly 230 percent. Lee and McCrary calibrate a version of the baseline dynamic model outlined in Section 3 and use the model to provide bounds for the elasticity of crime with respect to police and with respect to an expected sentence length. These bounds are nonparametric in the sense that they do not require a parametric restriction on the distribution of criminal benefits and instead require that the distribution of criminal benefits has a weakly declining density. As discussed by Viscusi (1986), for example, this assumption arises naturally from the fact that crime is largely a transfer from victim to offender, and that criminal opportunities worth more to an offender are likely to be taken out of circulation, or possibly "hardened" in some way, so as to reduce the frequency of such opportunities.

Jacob, Lefgren, and Moretti (2007) emphasize another important dynamic aspect of criminal behavior: intertemporal substitution. Using data from the National Incident Based Reporting System (NIBRS), these authors document that, (1) conditional on weather conditions in period t and other period t controls, weather conditions in period $t-1$ are strongly correlated with crime in period $t-1$, and that (2) using weather in period $t-1$ as

an excluded instrument, instrumental variables estimates of the elasticity of the crime rate in one week with respect to the crime rate in the past week is about -0.22 for violent crime and -0.17 for property crime. The elasticity with respect to two weeks prior is -0.16 for violent crime and -0.14 for property crime, and the elasticity seems to decline at longer lags.

The extent to which criminal labor supply may be substitutable, across time but also across space, is one of the great unsolved problems of crime control. If displacement is important, then costly attempts at saturation policing, known as "hot spots policing", may well simply shuffle crime from place to place or from time period to time period, with little overall impact on the present discounted value of national crime. This is an important consideration in light of the consensus within criminology that saturation policing is an effective tactic (Braga 2005). Perhaps prompted by consultant criminologists, many police departments have adopted these tactics in recent years (Heinzmann 2004, Raghavan 2005, Eiserer 2005, Higgins 2006, Hoover 2007, Katz 2009, Hunt 2009). Displacement is not widely discussed in the literature, perhaps in part because it is so hard to measure. The "catchment area" for the crime displaced from any hot spot is hard to identify *a priori*, making it hard to design a research study to quantify the importance of displacement.

The Jacob, Lefgren, and Moretti (2007) findings provide evidence that displacement is an important phenomenon, at least within the time dimension. In light of the Lee and McCrary (2009) findings, one possible inter-

pretation is that it is easier to detect displacement in time than in space, because offenders have a taste for the present and would not want to defer activity much beyond the period during it was suppressed.[4] However, Jacob, Lefgren, and Moretti emphasize that their results are difficult to reconcile with extreme scenarios in which either (1) potential offenders are unable or unwilling to save or borrow, or (2) potential offenders have long time horizons and access to perfect capital markets. In the first case, potential offenders solve a static problem and a period $t-1$ income shock cannot affect period t behavior. In the second case, by the permanent income hypothesis, the only role for a period $t-1$ income shock to affect period t behavior is through income effects, which the authors note are generally thought of as small.[5] Consequently, the authors prefer a hybrid scenario, in which offenders have a short time horizon, with no saving or borrowing. Because of the short time horizon, the scope for income effects is substantial. The authors note that they find much larger displacement effects for property crimes such as car theft that are associated with large income gains.

While the Lee and McCrary (2005, 2009) and Jacob, Lefgren, and Moretti (2007) studies highlight that offenders may have short time horizons, this conclusion is far from settled in the literature. A series of important dynamic articles in the *International Economic Review* (*IER*) in 2004 each assume

[4]On prior grounds, displacement in space seems likely to be much more diffuse than displacement in time, particularly with respect to activities such as drug transactions.

[5]In addition, income effects do not predict the "fadeout" pattern in the elasticities discussed above and thus are not a convincing explanation for the findings.

long time horizons for potential offenders. These articles emphasize a variety of dynamic mechanisms that are relevant for the study of crime.

Huang, Laing, and Wang (2004) present a nuanced model with endogenous human capital accumulation, heterogeneous firms, and labor market search. A major feature of the article is a clear discussion of equilibrium. A limitation of the article, however, is the assumption that punishment is experienced as a one-period utility loss, much as in the Becker (1968) model. This limits the ability of the model to accomodate non-responsiveness of offenders to prison due to short time horizons. The article makes no effort at either estimation or calibration.

Imai and Krishna (2004) emphasize the idea that engaging in crime today may have negative consequences for completion of education and for employment and wages in the future. The approach is a partial equilibrium dynamic structural model, along the lines of that discussed in Lee and McCrary (2009), but with a much richer specification of potentially heterogeneous preferences, both between persons and across time. The paper explicitly estimates the dynamic model using data from the 1958 Philadelphia Birth Cohort Study.[6]

The major empirical conclusion of the article is that the prospect of future reductions in labor market employability and remuneration leads to reduced current period criminal activity ("dynamic deterrence"). However, this substantive conclusion is intrinsically linked with the substantive conclusion of long time horizons: Imai and Krishna (2004) estimate an annual discount

[6]These data are not often used within economics, but are available from ICPSR.

factor of 0.99. Further, this estimate is based on the assumption that punishment lasts one period. Intuitively, if the marginal offender has a short time horizon, then this limits the scope for dynamic deterrence, just as it limits the scope of the effectiveness of prison as a punishment.

Another important article in the same issue of the *IER* is Imrohoroglu, Merlo, and Rupert (2004), which seeks to understand the extent to which the crime drop of the 1990s is consistent with a dynamic model with heterogeneous agents. The model is calibrated to match the 1980 crime rate. The inputs to this calibration are the apprehension probability (based on Uniform Crime Reports (UCR) data), the mean and standard deviation of predicted log real wages (based on Current Population Survey (CPS) data), and finally the age distribution, education distribution, and unemployment rate (CPS). In a departure from the analysis of Imai and Krishna (2004) and Huang, Laing, and Wang (2004), Imrohoroglu, Merlo, and Rupert (2004) allow punishment to be of varying durations. Perhaps importantly, the calibration exercise assumes an annual discount factor of 0.989.

A major focus of this article is the capacity of the calibrated model to match the 1996 crime rate of 4.6 percent. The model performs remarkably well, predicting a 1996 crime rate of 4.7 percent. The article provides a decomposition of the contribution of the different inputs to this conclusion, analogous to standard decomposition exercises such as Blinder (1973), Oaxaca (1973), or DiNardo, Fortin, and Lemieux (1996). Probably the most remarkable aspect of this decomposition is that a *ceteris paribus* increase in

the clearance rate from its 1980 level of 16.8 to its 1996 level of 18.5 percent is predicted to have decreased the crime rate from 5.6 to 3.2 percent. This implies an elasticity of crime with respect to police of -4.3, several orders of magnitude larger than those discussed in the quasi-experimental literature (e.g., Di Tella and Schargrodsky 2004). Another interesting aspect of the decomposition exercise is that the 20 percent increase in the standard deviation of log income is predicted to have increased crime by 59 percent, implying an elasticity of 2.9.

Burdett, Lagos, and Wright (2003, 2004) present a search equilibrium framework in which crime, unemployment, and wage inequality are interrelated phenomena.[7] Punishment in their model lasts multiple periods, but the focus is not on the magnitude of deterrence elasticities. Instead, these papers emphasize that the introduction of crime to search equilibrium models leads to wage dispersion, non-monotonicity between some policy parameters and crime, and multiple equilibria. This latter conclusion is consistent with the empirical findings of Glaeser, Sacerdote, and Scheinkman (1996), as these authors emphasize. A particularly provocative finding from calibration results is that social support programs can lead to more crime, rather than less, because of the need to raise taxes to pay for them. Higher taxes discourage work and encourage crime in the model.

Like most of the dynamic papers in the literature involving numerical

[7] Burdett, Lagos, and Wright (2004) differs from Burdett, Lagos, and Wright (2003) in that it incorporates on-the-job search and presents calibration results and counterfactual policy evaluations.

results, Burdett, Lagos, and Wright (2004) assume a long time horizon for potential offenders. Both firms and individuals are assumed to have an annual discount factor arbitrarily close to 1. It is not clear from the discussion how much the qualitative conclusions of the model are affected by this calibration choice.

The remaining dynamic article in the *IER* special issue is Lochner (2004), which focuses on the process of human capital accumulation and its implications for crime. A particular focus is to combine an analysis of individual schooling decisions with crime decisions. This focus is used to shed light on the age-crime profile and on the stark differences in criminal involvement between those with little and much education.

An important qualitative conclusion from Lochner's (2004) analysis is that the short- and long-run crime return to government investments may be quite different. Generally, long-run crime reductions will be much larger than short-run crime reductions. For example, a government subsidy to stay in school or enroll in a job training program may reduce crime in the short-run, by shifting prices. To the extent that these human capital investments increase future legitimate labor market wages, a beneficiary of the government subsidy may have reduced criminal involvement, even after graduating from school or ending training. Consequently, the long-run return to government investment can be much larger than the short-run return. These kinds of considerations are strengthened by the possibility of criminal, as well as legitimate, human capital accumulation.

The most recent contribution to the dynamic literature on crime is Sickles and Williams (2008). These authors argue that the literature has focused too much on deterrence elasticities and not enough on social programs that might foster what is termed "social capital". Social capital is conceptualized as a stock, subject to depreciation, which discourages participation in crime; the stock is increased when the individual abstains from crime and pays dividends in terms of utility directly as well as earnings. In this formulation, social capital is a mechanism for persistence in choices. The authors propose a model in which individuals devote time to legitimate work, leisure, and crime. As with many of the papers in this literature, Sickles and Williams (2008) assume that punishment is experienced in one period. Moreover, offenders are assumed to have long time horizons, with annual discount factors set to 0.95.

3 Offenders

In this section, I lay out a simple dynamic model of behavior. I first describe in detail a baseline model with a representative agent and time homogeneity (or stationarity), discussed in Lee and McCrary (2009). Then, I sketch a slightly more general model and show how to construct the likelihood function for either model using longitudinal data on arrests. This kind of data is available from nearly every state government and some data along these lines are publicly available.

3.1 Baseline Model

Suppose that the agent faces the same problem throughout daily life: each day, a criminal opportunity presents itself, and the opportunity may or may not be worth taking advantage of. Let the criminal benefit in any given period and state be denoted B, viewed as random draw from a distribution with distribution function $F(b)$ and density function $f(b)$. The agent lives for an infinite number of periods, which is a reasonable approximation when each period is taken to be a day. Thinking of crimes arriving through some stochastic process at a rate of one a day is also a reasonable description of timing, in light of the typical frequency of criminal involvement documented in the literature (Cohen 1986).

If the agent commits crime, he runs the risk of apprehension, which occurs with probability p. If apprehended, the agent is immediately detained for S periods, where S is random and can take on values $1, 2, 3, \ldots$. Let π_s denote the probabibility that $S = s$. The only quantities that are allowed to be stochastic are the value of the criminal benefit, the event of apprehension, and the sentence length conditional on apprehension. Note that S refers both to a "sentence", as per an adjudication, as well as to pre-trial detention such as bail or even questioning by police. It is intended to represent any period of time the agent is unable to engage in another crime, by virtue of having been apprehended for a given crime. Hence a quantity like $E[S]$ refers to the expected sentence, conditional on arrest, and thus folds in the probability of going to trial, the probability of conviction conditional on going to trial, and

so forth.

While detained, the agent cannot commit crimes and receives flow utility $a - c$. If the agent is free and abstains from crime, he receives flow utility a. If the agent is free and commits crime without being apprehended, he receives flow utility $a + B$. Apprehension occurs immediately or never, so if the agent elects to commit crime and is apprehended, he does not receive the criminal benefit B. To make the problem non-trivial, we assume that $a + b > a > a - c$ for almost every b in the support of B. For the sake of simplicity, we take a and c to be deterministic.

In the baseline model, these three flow utilities are assumed to be constant in time and across states, as are the probability of apprehension and the distribution of sentence lengths. This simplifies many of the calculations involved in the model. Although a can be normalized to 0 in the baseline model, due to time homogeneity and expected utility maximization, we retain it in the expressions that follow to simplify the connection to the time heterogeneous model described below.

In each period t, the agent chooses a strategy—an action for the current period and a set of contingent plans for actions in subsequent periods—seeking to maximize his expected present discounted value, or $E_t \left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} u_t\right]$, where E_t is the expectation operator conditional on information available as of period t, δ is the discount factor, and u_t is either $a - c$, a, or $a + B$, depending on the agent's choices, whether the criminal opportunity materializes, whether he has been apprehended for any crimes committed, and

whether he is currently detained. In the baseline model time homogeneity is assumed. Hence between periods t and $\tau > t$, no additional information is obtained by the agent, and we simply write $E[\cdot]$ in place of $E_t[\cdot]$.

An agent electing to engage in a crime of value $B = b$ receives payoff $a + b + \delta E\left[V(B)\right]$ if he gets away with it, where $V(B)$ is the value of being free and having received criminal opportunity B. The value of being free to commit crime is stochastic, because at the time the agent is deciding whether to engage in crime, the current value of crime, b, is known, but the value of crime for next period, B, is not yet known. By time homogeneity, the value of the criminal opportunity is constant in time, so is not subscripted by time.

An agent electing to engage in a crime of value $B = b$ who is apprehended receives, with probability π_s, the sentence length s and hence the payoff

$$(a - c)\left(1 + \delta + \delta^2 + \cdots + \delta^{s-1}\right) + \delta^s E\left[V(B)\right] \tag{1}$$

An agent who abstains from crime receives payoff $a + \delta E\left[V(B)\right]$. Thus, the value of being free and receiving criminal opportunity $B = b$ is

$$\begin{aligned} V(b) \quad = \quad & \max\Bigg\{ a + \delta E[V(B)] \quad , \\ & p\sum_{s=1}^{\infty}\pi_s\left[(a-c)\frac{1-\delta^s}{1-\delta} + \delta^s E[V(B)]\right] + (1-p)\Big[a + b + \delta E[V(B)]\Big]\Bigg\} \end{aligned} \tag{2}$$

The optimal strategy for the agent is a "reservation policy" whereby there exists a reservation criminal benefit, b^*, with the property that the agent

faced with a criminal opportunity such that $B > b^*$ will elect to commit crime, the agent faced with a criminal opportunity such that $B < b^*$ will elect to abstain, and the agent faced with a criminal opportunity such that $B = b^*$ will be indifferent between these options.[8] Indifference implies that if $B = b^*$, the agent gets the same value from committing crime as from abstaining. Formally, this implies

$$\begin{aligned} a + \delta E\left[V(B)\right] &= p\sum_{s=1}^{\infty}\pi_s\left[(a-c)\frac{1-\delta^s}{1-\delta} + \delta^s E\left[V(B)\right]\right] \\ &+ (1-p)\left[a + b^* + \delta E\left[V(B)\right]\right] \end{aligned} \tag{3}$$

Rearranging, we have

$$b^* = c\frac{p}{1-p}\left[1 + \sum_{s=1}^{\infty}\pi_s\frac{\delta-\delta^s}{1-\delta}\left(1 + \frac{(1-\delta)E\left[V(B)\right] - a}{c}\right)\right] \tag{4}$$

When $\delta = 0$, the potential offender disregards the future and this reservation benefit reduces to $cp/(1-p)$, or the reservation benefit given by the static Becker model. The same simplification obtains when there is no chance of being detained longer than one period, i.e., when $\pi_s = 0$ for $s = 2, 3, \ldots$.

By time homogeneity, $E\left[V(B)\right]$ is constant in time and hence can be

[8] This mimics the standard "reservation wage" property of a job search model. See, for example, McCall (1970) or the textbook treatments in Adda and Cooper (2003) and Ljungqvist and Sargent (2004).

calculated from the following recursive relationship:

$$\begin{aligned} E\left[V(B)\right] &= F(b^*)\Big[a+\delta E\left[V(B)\right]\Big] \\ &+ (1-F(b^*))(1-p)\Big[a+E[B|B>b^*]+\delta E\left[V(B)\right]\Big] \\ &+ (1-F(b^*))p\sum_{s=1}^{\infty}\pi_s\left[(a-c)\frac{1-\delta^s}{1-\delta}+\delta^s E\left[V(B)\right]\right] \end{aligned} \tag{5}$$

Intuively, this equation states that in expectation, the value of being free consists of three distinct pieces: (i) the value associated with drawing a criminal opportunity that is not worth committing, (ii) the value associated with drawing a criminal opportunity that is worth committing and for which one is not apprehended, and (iii) the value associated with drawing a criminal opportunity that is worth committing and for which one is apprehended. Each of these three pieces depends in turn on the value of being free. One can rearrange this recursion and solve for $E[V(B)]$, but a a more intuitive expression is obtained by eliminating the infinite sum from equation (5) using the indifference result in equation (3).[9] This yields

$$(1-\delta)E\left[V(B)\right] = a+(1-F(b^*))(1-p)E[B-b^*|B-b^*>0] \tag{6}$$

Equation (6) shows that the annuitized value of being free and obtaining a

[9]Some algebra is avoided by instead recognizing that by indifference and continuity of $V(\cdot)$, $V(b)=a+\delta E[V(B)]+(1-p)(b-b^*)\mathbf{1}(b>b^*)$ where $\mathbf{1}(\cdot)$ is 1 if the event in parentheses is true and is 0 otherwise. Hence $E[V(B)]=F(b^*)\left\{a+\delta E[V(B)]\right\}+(1-F(b^*))\left\{a+\delta E[V(B)]+(1-p)E[B-b^*|B>b^*]\right\}$, which implies equation (6).

draw from the criminal opportunity distribution is equal to the flow utility associated with abstaining, plus the option value of crime. The option value is equal to the probability of getting a draw that is worth taking advantage of and avoiding apprehension, times the relative expected value of a criminal opportunity deemed worth taking advantage of.[10]

Although b^* is only implicitly defined in equation (4), it is not difficult to compute. Define $\nu = \sum_{s=1}^{\infty} \pi_s \frac{\delta - \delta^s}{1-\delta}$ and the functions[11]

$$\psi(b) = (1 - F(b))E\left[B - b | B - b > 0\right] = \int_b^{\infty} (1 - F(t))dt \tag{7}$$

$$\tilde{\psi}(b) = -F(b)E\left[B - b | B - b < 0\right] = \int_0^b F(t)dt \tag{8}$$

Then, combining equations (4) and (6), we have

$$b^* = c\frac{p}{1-p}\left[1 + \nu\left(1 + \frac{(1-p)\psi(b^*)}{c}\right)\right] = \alpha + \beta\psi(b^*) \tag{9}$$

$$= \frac{\alpha + \beta E[B]}{1+\beta} + \frac{\beta}{1+\beta}\tilde{\psi}(b^*) \equiv \tilde{\alpha} + \tilde{\beta}\tilde{\psi}(b^*) \tag{10}$$

where $\alpha = c\frac{p}{1-p}(1+\nu) > 0$, $\beta = p\nu > 0$, and where equation (10) follows from the fact that $\psi(b) = E[B] - b + \tilde{\psi}(b)$.[12] Equations (9) and (10) can be

[10] Throughout this article, I refer to equations like (3) and (4) as pertaining to *indifference* and to equations like (5) and (6) as pertaining to *rational expectations*.

[11] Note that (1) ν is guaranteed to be between 0 and $\min\{E[S]-1, \delta/(1-\delta)\}$, is increasing in the severity of sentences, can be rewritten as $\sum_{s=1}^{\infty} \delta^s P(S > s)$, and has derivative $\frac{\partial \nu}{\partial \delta} = \sum_{s=1}^{\infty} s\delta^{s-1}P(S > s) > 0$; (2) $\psi(\cdot)$ is a positive, decreasing, convex function with $\psi'(b) = -(1 - F(b))$, $\psi''(b) = f(b)$, $\psi(0) = E[B]$, and $\lim_{b\to\infty} \psi(b) = 0$; and (3) $\tilde{\psi}(\cdot)$ is a positive, increasing, convex function with $\tilde{\psi}'(b) = F(b)$, $\tilde{\psi}''(b) = f(b)$, $\tilde{\psi}(0) = 0$, and $\lim_{b\to\infty} \tilde{\psi}(b) = \infty$.

[12] The fact that $\psi(b) = E[B] - b + \tilde{\psi}(b)$ can be seen by adding and subtracting $\int_0^b (1 -$

used to demonstrate that b^* is unique, and equation (10) demonstrates the viability of solving for b^* using a contraction mapping, since $\tilde{\beta} < 1$. That is, one can apply equation (10) repeatedly, starting at any initial guess for b^*. Such a process is guaranteed to converge; to obtain a solution within any given tolerance of b^* requires only a finite number of steps; and usually a double precision solution can be obtained in a modest number (e.g., less than 20) of steps. It is easy to implement such a procedure in standard software packages as part of a routine for evaluating a likelihood function or a moment function.[13]

FIGURE 1 ABOUT HERE

Figure 1 presents a graphical analysis of equations (9) and (10) for a particular parameterization of the baseline model. In both panels, the x-axis is a possible value of the reservation benefit, or b, and the solid line is the 45 degree line. In panel A, the dashed curve is $\alpha+\beta\psi(b)$, which is monotonically decreasing with a positive intercept for any criminal benefit distribution, and hence guaranteed to cross the 45 degree line once and only once. In panel B, the dashed curve is $\tilde{\alpha} + \tilde{\beta}\tilde{\psi}(b)$, which is monotonically increasing with a positive intercept. The uniqueness of the intersection then follows since the derivative of $\tilde{\alpha} + \tilde{\beta}\tilde{\psi}(b)$ is strictly less than one. In both figures, the vertical

$F(t))dt$ from the definition for $\psi(b)$.

[13] A slightly faster approach is to apply Newton's method to either equation (9) or (10). For a detailed discussion of computational issues in dynamic models from a more general perspective, see Rust (1996).

dotted line indicates the value of the reservation benefit, b^*.[14]

The ex ante probability of crime commission among those not detained is given by $G(b^*) \equiv 1 - F(b^*)$. Thus, factors that work to increase b^* work to decrease crime on the part of the undetained population. I now turn to comparative statics regarding the crime rate. Crime is unambiguously reduced by increases in: (1) δ, the discount factor; (2) p, the probability of apprehension; and (3) c, the per period relative cost of prison. This is because increasing any of these quantities increases α, which shifts out the dashed curve $\alpha + \beta\psi(b)$. Increasing δ and p also increases β, which results in a further shift out in $\alpha + \beta\psi(b)$. Any shift out in this curve increases b^* and hence decreases $G(b^*)$, i.e., lowers the ex ante probability of crime on the part of an undetained person.[15] The remaining comparative statics I consider pertain to distributions: that of sentence lengths, or $\{\pi_s\}_{s=1}^{\infty}$, and that of the criminal benefit, or $F(\cdot)$. Regarding sentence lengths, we have the following result: as long as $\delta > 0$, i.e., as long as the agent cares about the future, any rightward (i.e., first order stochastic dominant) shift in the distribution of sentence lengths unambiguously reduces crime. The easiest way to recognize this is to note that any rightward shift in the distribution of sentence lengths increases $P(S > s)$ for at least one s. Since ν can be

[14]This formulation shows

$$\frac{c\frac{p}{1-p}(1+\nu) + p\nu E[B]}{1+p\nu} \quad < \quad b^* \quad < \quad c\frac{p}{1-p}(1+\nu) + p\nu E[B]$$

which holds regardless of the criminal benefit distribution. These bounds are tight when $p\nu$ is small.

[15]These results can also be established rigorously using the implicit function theorem.

written as $\nu = \sum_{s=1}^{\infty} \delta^s P(S > s)$, such a shift increases ν, which increases β, shifting out $\alpha + \beta\psi(b)$. Of course, when $\delta = 0$, ν is exactly zero and does not respond to changes to the distribution of sentences.

Regarding the distribution of criminal benefits, nothing general can be said about the effect of a shift to the right in the distribution on the crime rate. This counterintuitive result ("Shouldn't there always be more crime when crime is worth more?") is a distinctive feature of a dynamic model of crime. The reason for the ambiguity of such a comparative static is that even though a shift to the right increases b^* unambiguously, by definition the survivor curve $1 - F(b)$ also shifts out. In some settings, the shift out in the survivor function can fully offset the shift out in the reservation benefit, resulting in a net reduction in crime.

To explain this issue, consider two examples of shifts to the right in the criminal benefit distribution. The first example is associated with an increase in crime. For $k > 1$, the survivor function $1 - F(b)$ increases for values $b > kb^*$, but is the same for $b \leq kb^*$. This leads to an increase in the reservation benefit because $\psi(\cdot)$ mechanically increases (see equation (7)). However, since the survivor function does not shift in the neighborhood of b^*, the survivor function is unaffected at the new b^*, and the net effect is an increase in crime.

The second example is instead associated with a decrease in crime. The initial criminal benefit distribution has support on $[0, 1]$ and further has a mass point at 1. Let q denote the probability that the benefit is exactly 1.

Now consider a rightward shift in the distribution such that the mass point is increased from 1 to $\overline{B}$. Such a move necessarily increases b^*, and for large enough $\overline{B}$ moves b^* above 1. In this example, whenever b^* moves above 1, the probability of crime on the part of the free is immediately q, which can be arbitrarily small and hence lower than any original probability of crime.

The intuition behind these two examples is as follows. When the criminal benefit distribution shifts to the right, two conceptually different effects impinge on behavior: the *current wage effect* and the *opportunity cost effect*. The current wage effect is that the value of crime is higher, which would lead to more crime if $E[V(B)]$ were held constant. The opportunity cost effect is that *all* future draws are likely to be better than they otherwise would be: that is, $E[V(B)]$ is higher. Hence, committing crime next period puts the agent at risk of being imprisoned and hence unable to avail himself of criminal opportunities two periods hence, three periods hence, and so on. The opportunity cost effect tends to reduce crime. In the extreme, the opportunity cost effect can dominate. Intuitively, we can imagine a shift in the distribution of criminal benefits such that there is an outside chance at riches so fantastic that the agent chooses to spend the vast majority of his life abstaining from crime, hoping to be free and able to avail himself of a criminal opportunity so rare that it almost never arrives. There exists a class of criminal benefit distributions for which the current wage effect dominates the opportunity cost effect, but characterizations of this class have not yet been studied in the literature.

For the purposes of this chapter, the most interesting policy questions to which the model speaks pertain to (1) offender responses to the threat of apprehension (p) and (2) offender responses to the threat of punishment, conditional on apprehension ($E[S]$). We have

$$\eta_p \equiv \frac{\partial G(b^*)}{\partial p}\frac{p}{G(b^*)} = -\frac{f(b^*)}{G(b^*)}\left[b^* + c\left(\frac{p}{1-p}\right)^2(1+\nu)\right]\frac{1}{1+\nu p G(b^*)} \tag{11}$$

$$\eta_{E[S]} \equiv \frac{\partial G(b^*)}{\partial E[S]}\frac{E[S]}{G(b^*)} = -\frac{f(b^*)}{G(b^*)}\left[b^* - c\frac{p}{1-p}\right]\eta_\nu\frac{1}{1+\nu p G(b^*)} \tag{12}$$

where $0 < \eta_\nu < \frac{E[S]}{E[S]-1}$ is the elasticity of ν with respect to $E[S]$. These expressions consist of three factors which can be loosely characterized as follows. The first factor, $f(b^*)/G(b^*)$, depends on the shape of the density of criminal benefits local to the reservation benefit. This modulates the prevalence of individuals on the margin of the crime participation decision. The second factor, given in brackets, pertains to the extent to which the reservation benefit is changed by the policy parameters p and $E[S]$.[16] The third and final factor, $1/\left(1+\nu p G(b^*)\right)$, reflects the opportunity cost of crime. As discussed in Lee and McCrary (2009), this factor reflects the fact that when p or $E[S]$ increases, crime in the future is less attractive (the opportunity cost effect). This reduces the anticipated criminal involvement in the future, which reduces $E[V(B)]$ and thus leads to a lessening of the opportunity cost of imprisonment. By analogy with consumer theory, one could

[16]For the sentencing elasticity, the additional term η_ν reflects the fact that $E[S]$ only affects crime through its impact on ν.

imagine a temporary shift in p or $E[S]$ that held $E[V(B)]$ constant. Since $1\big/\left(1+\nu pG(b^*)\right) < 1$, the effect of such a policy reform on behavior would be strictly larger than the effect of a policy reform that shifted p or $E[S]$ for all time.

Finally, equations (11) and (12) clarify that in this model, increasing p almost always has a bigger deterrence effect than increasing $E[S]$. In particular, we have

$$\frac{\eta_p}{\eta_{E[S]}} = \frac{b^* - c\frac{p}{1-p} + c\frac{p}{(1-p)^2}(1+p\nu)}{b^* - c\frac{p}{1-p}}\frac{1}{\eta_\nu} \tag{13}$$

The first term, $\frac{b^*-c\frac{p}{1-p}+c\frac{p}{(1-p)^2}(1+p\nu)}{b^*-c\frac{p}{1-p}}$, always exceeds 1, and the second term, $\frac{1}{\eta_\nu}$, exceeds 1 except for very patient individuals facing very short expected sentence lengths. Intuitively, punishment occurs in the future and apprehension occurs in the present. Thus a change to expected punishments imposes smaller costs than a change to the probability of apprehension. Taking limits as $\delta \to 0$, we see that for individuals with arbitrarily short time horizons, policing becomes infinitely more effective than punishment.[17]

[17] To a certain extent, this result can be viewed as preordained by our timing convention that detention is simultaneous with apprehension. When detention occurs the period after apprehension, and we take limits as $\delta \to 0$, both policing and punishment are ineffective. However, in continuous time, we again obtain the result that policing is infinitely more effective than punishment for individuals with arbitrarily short time horizons.

3.2 Heterogeneity

Assumptions of time homogeneity are restrictive for most crime applications, due to the rapid changes in criminal involvement as youths move into adulthood (Lochner 2004). Similarly, the literature has emphasized the cross-sectional heterogeneity in criminal propensity, with a small number of individuals committing a great number of crimes (Visher 1986, Piehl and DiIulio 1995, Blumstein 2002).

When the flow utilities vary over time and across persons, the model is similar, but the notation becomes more complicated, and careful attention must be paid to the nature of conditional independence assumptions which are invoked.

If agent i elects in period t to engage in a crime of value $B_{it} = b_{it}$ and gets away with it, he receives payoff $a_{it} + b_{it} + E_{it}[V_{i,t+1}(B_{i,t+1})]$, where $E_{it}[\cdot]$ is the expectation operator conditional on information regarding agent i available at period t.[18] If the agent is instead apprehended, he receives, with probability π_{its}, the sentence s and the payoff

$$\begin{aligned}(a_{it} - c_{it}) + \delta(a_{i,t+1} - c_{i,t+1}) + \cdots + \delta^{s-1}(a_{i,t+s-1} - c_{i,t+s-1}) \\ +\delta^s E_{it}[V_{i,t+s}(B_{i,t+s})] \end{aligned} \tag{14}$$

If the agent abstains from crime, he receives the payoff $a_{it}+\delta E_{it}[V_{i,t+1}(B_{i,t+1})]$.

[18] We continue to assume that a_{it} and c_{it} are deterministic. They may evolve over time in a predictable fashion, however. Generally, if there are state variables, then $E_{it}[\cdot]$ is an expectation conditional on those state variables, as of period t.

Thus, the value to agent i of being free in period t and receiving criminal opportunity $B_{it} = b$ is

$$
\begin{aligned}
V_{it}(b) &= \max\Bigg\{ a_{it} + \delta E_{it}[V_{i,t+1}(B_{i,t+1})] \ , \qquad (15) \\
&\qquad p_{it} \sum_{s=1}^{\infty} \pi_{its} E_{it} \left[\sum_{j=0}^{s-1} \delta^j (a_{i,t+j} - c_{i,t+j}) + \delta^s V_{i,t+s}(B_{i,t+s}) \right] \\
&\qquad +(1-p_{it})\Big(a_{it} + b + \delta E_{it}\left[V_{i,t+1}(B_{i,t+1})\right] \Big) \Bigg\} \\
&= a_{it} + \delta E_{it}\left[V_{i,t+1}(B_{i,t+1})\right] + (1-p_{it})(b - b^*_{it})\mathbf{1}(b > b^*_{it}) \qquad (16)
\end{aligned}
$$

where the simplification in equation (16) occurs by the same reasoning described for the baseline model, and where the reservation benefit is given by

$$
\begin{aligned}
b^*_{it} &= \frac{p_{it}}{1-p_{it}} \Bigg\{ a_{it} + \delta E_{it}\left[V_{i,t+1}(B_{i,t+1})\right] \qquad (17) \\
&\qquad - \sum_{s=1}^{\infty} \pi_{its} E_{it} \left[\sum_{j=0}^{s-1} \delta^j (a_{i,t+j} - c_{i,t+j}) + \delta^s V_{i,t+s}(B_{i,t+s}) \right] \Bigg\}
\end{aligned}
$$

The indifference equation (17) indicates that b^*_{it} depends on expectations of future utility flows and values of the dynamic program. From the rational expectations equation (16), we see that

$$
E_{i,t-1}[V_{it}(B_{it}) - \delta V_{i,t+1}(B_{i,t+1})] = a_{it} + E_{i,t-1}[(1-p_{it})(B_{it} - b^*_{it})\mathbf{1}(B_{it} > b^*_{it})] \qquad (18)
$$

As in the baseline model, the indifference and rational expectations equa-

tions (17) and (18) are the main equations of this model. Precisely how to form the relevant expectations differs substantially across variations of this model. For these models to be tractable, sufficient structure has to be placed on flow utilities, system variables, and laws of motion that equations (17) and (18) can be solved. The literature is sufficiently in its infancy that there has not yet emerged a "workhorse" model, so assumptions differ.

However, it is straightforward to explain the basic idea. Suppose "eventual homogeneity"—that is, suppose that at some age T, the individual faces a time homogenous environment. Under this assumption, the baseline model can then be used to solve for the $b^*_{i,T}$ and $E_{i,T-1}[V_{i,T}(B_{i,T})]$. By eventual homogeneity, we have $E_{i,T-1}[V_{i,T+j}(B_{i,T+j})] = E_{i,T-1}[V_{i,T}(B_{i,T})]$ for all $j > 0$. Then equation (17) can be used to solve for $b^*_{i,T-1}$, and (18) can be used along with model assumptions to backwards iterate to obtain the value function. Proceeding iteratively, one can obtain the reservation benefit for each agent for each period of time.

3.3 Connecting Theory to Data

We now discuss the difficult issue of how to connect the theory outlined above to data. This differs from the norm in other areas of economics in which structural modeling is used, because of the nature of the information available in crime applications. For example, in job search applications, it is standard to observe the wage for those who accept work. In crime settings, it is uncommon to observe the value of the criminal benefit, even among those

engaging in crime. Sometimes this is due to data availability, and other times it is because the value of the particular crime to the offender may be inherently difficult to quantify (e.g., assault).

The many varieties of possible data sets and models one could use means that it is difficult to give any general description of how to connect the theory to the data. I give one example, tailored to the most commonly available data: individual arrest histories ("rap sheets").[19]

The sequence b^*_{it} is a panel data set of predicted reservation benefits that depend on particular values of the vector of structural parameters, which I will denote θ. Precise details of what is included in θ differ from model to model, but a typical implementation would include δ, parameters pertaining to the flow utilities, the criminal benefit distribution, motion equations, and so on.[20]

Let $h_i(t|\theta) \equiv h_i(t) = p_{it}G_{it}(b^*_{it})$ denote the predicted probability of arrest from the model for agent i at time t, conditional on being free. Define the cumulated hazard, $H_i(t|\theta) \equiv H_i(t) = -\ln \mathcal{S}_i(t)$, where $\mathcal{S}_i(t) = \prod_{\tau=1}^{t}(1 - h_i(\tau))$ is the survivor function. If $D_{it} = H_{it}^{-1}(H_i(t-1)+\varepsilon)$, where $H_{it}^{-1}(v) \equiv \min\{t : H_i(t) \geq v\}$ and ε is distributed standard exponential, then D_{it} is a duration consistent with the hazard sequence $h_i(t|\theta), h_i(t+1|\theta), \ldots$ (Devroye

[19]It is typically difficult to obtain such arrest histories merged with information on prison stays. When merged arrest-prison data are available, the issues are slightly different than discussed below. We focus on the more challenging case of what to do when no information on prison stays is available.

[20]I assume that the sentence length distribution is modeled separately, using publicly available data on sentencing.

1986, Section VI.2). That is, D_{it} is a duration consistent with being at risk of failure starting in period t.[21]

Let X_i denote age at first arrest, and let Y_i^* denote age at second arrest. I use the notation Y_i^* because age at second arrest may be censored. Under the model, age at second arrest may be viewed as having been generated as

$$Y_i^* = H^{-1}(H(X_i + S_i - 1) + \varepsilon_i) \tag{19}$$

where X_i is age at first arrest, S_i is sentence length, and ε_i is distributed standard exponential and independent of X_i and S_i.

I next use this representation to derive the log-likelihood function for observed age at second arrest, conditional on age at first arrest. To do so, I first derive the distribution of (latent) age at second arrest and then apply standard results on likelihood functions under censoring. Fix $y > X_i$, both integers. The event $Y_i^* \leq y$ is the same as the event $H(X_i+S_i-1)+\varepsilon_i \leq H(y)$, and the event $Y_i^* = y$ is the same as the event $H(y-1) < H(X_i+S_i-1)+\varepsilon_i \leq H(y)$. This leads to expressions for the conditional distribution function and

[21] There is no content to ε being distributed standard exponential. Rather, this is simply the duration data analogue to the well-known result that $F(Z)$ is distributed standard uniform, if Z has distribution function $F(\cdot)$.

conditional probability function of Y_i^*,

$$F_{Y^*|X}(y|X_i) = P(S_i \leq y - X_i|X_i) - \sum_{s=1}^{y-X_i} P(S_i = s|X_i)\mathcal{S}(y)\big/\mathcal{S}(X_i + s - 1) \quad (20)$$

$$f_{Y^*|X}(y|X_i) = h(y) \sum_{s=1}^{y-X_i} P(S_i = s|X_i)\mathcal{S}(y-1)\big/\mathcal{S}(X_i + s - 1) \quad (21)$$

One may verify that equation (21) is the first difference of equation (20). Both expressions can be calculated exactly given a known conditional distribution for sentences given age at first arrest.

Observed age at second arrest is a censored version of latent age at second arrest, i.e., $Y_i = \min\{Y_i^*, C_i\}$. This poses little difficulty once we have specified the distribution of the latent variable, however (Wooldridge 2002, Lawless 2003). Let $\kappa_i = 1$ indicate that the observation is censored and $\kappa_i = 0$ indicate that it is not. Then the log-likelihood function is

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left\{ (1 - \kappa_i) \ln f_{Y^*|X}(Y_i|X_i) + \kappa_i \ln \left(1 - F_{Y^*|X}(C_i|X_i)\right) \right\} \quad (22)$$

Most software packages allow the user to pass a function of data and parameters to a maximization routine that determines the parameter values that maximize the function, given the data. Passing $L(\theta)$ to such a maximization routine is thus a straightforward approach that yields maximum likelihood estimates of θ. The only computational problems with such an approach are that, depending on the scope of the data set and the model, it can be slow to evaluate the function $L(\theta)$. Note that analytical derivatives

of the likelihood function are extremely tedious to compute and likely not worth the effort, except in special cases. Consequently, the user will likely find it optimal to use numerical derivatives, which means many, many more function evaluations will be required before the routine has climbed to the top of the likelihood function.[22]

While I have discussed estimation by maximum likelihood in this subsection, this is certainly not the only available method for connecting the theory to the data. A leading technique is to choose a set of empirical quantities which are implicated by the structural parameters of interest. As long as there are a sufficient number of moments, one can hope to judiciously vary the structural parameters to match those moments.[23]

There is of course art in choosing which moments to match. In some settings, the research design may suggest the set of moments that should be matched. For example, Lee and McCrary (2009) use the change in offense rates around the 18th birthday to generate quasi-experimental moments.

However, even if the research design does not suggest which moments to match, there may be a transparency benefit associated with being able to name the source of identifying information in the approach. This transparency may be particularly valuable for these types of models. Obtaining

[22] A further consideration is the impact of computer precision on these function evaluations. For an introductory discussion, see Judd (1998). In my experience, these issues are particularly relevant for approximating infinite sums.

[23] An interesting technical issue that can arise in this context is that because of the nonlinearity of dynamic model, some moments that can be observed may not within the range of the model. This can often be remedied by minor alterations to the model, but it is not always obvious how the model needs to be altered.

and interpreting the score in dynamic crime models, even for the simplest versions, is costly; although it can be done, I am aware of no results in the literature along these lines.

4 Government

In this section, I consider the government's problem instead of the offender's problem. Government decides what level of resources should be devoted to fighting crime and how to use those resources within the criminal justice system. Expenditures on police and other uses are related to the "system parameters" discussed in Section 3. When government hires additional police officers, it does so with the aim of increasing the probability of apprehension, or p. When arrests per officer are only negligibly affected by the number of officers, then a 5 percent increase in the number of officers is associated with a 5 percent increase in p. When government passes laws for sentence enhancements, abolishment of parole, mandatory minimum sentences, and so on, it shifts the distribution of detention times for an arrestee to the right. Thus, a reasonable approximation to the problem facing government is minimization of the present discounted value of the crime burden by judicious choice of p_t and $P(S_t \geq s)$, subject to an intertemporal budget constraint.

This suggests the following formalization of the government's problem:

$$\min_{\{p_t, P(S_t \geq s)\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t C_t \right] \tag{23}$$
$$\text{s.t.} \quad A_{t+1} = R_{t+1}(A_t + I_t - B_t) \quad \text{and} \quad A_t \geq \underline{A}$$

Here, p_t is the aggregate apprehension probability, $P(S_t \geq s)$ is the aggregate survivor function for a sentence length, β summarizes the government's taste for the present, C_t is the probability that a person selected at random at time t from among the entire population (both free and in prison) is engaged in crime, A_t is criminal justice "assets", $\underline{A}$ is a minimum level of assets below which government cannot go, I_t is criminal justice revenues, B_t is per capita criminal justice expenditures, and R_{t+1} is the gross return to assets between periods t and $t+1$. Below, we will make use of the notation Q_t, or the probability that a person selected at random at time t is in prison. Although C_t and Q_t are probabilities, it is useful to think of these terms as capturing crime per capita and prisoners per capita, respectively. The government takes I_t and R_t to be exogenous; all other quantities—in particular C_t, Q_t, and B_t—are endogenous to the choice variables p_t and $P(S_t \geq s)$.

Both crime per capita and criminal justice expenditures per capita are related to prisoners per capita. Since prisoners per capita is a stock, the flow rate into and out of prison dictates the level. These considerations lead to

the restrictions

$$C_t = (1 - Q_t)G_t \tag{24}$$

$$B_t = w_t p_t + r_t Q_t \tag{25}$$

$$Q_t = (1 - Q_{t-1})G_{t-1}p_{t-1} + Q_{t-1}(1 - X_{t-1}) \tag{26}$$

where G_t is the probability of crime on the part of someone *not* in prison, $w_t p_t$ is the per capita apprehension ("policing") budget, $r_t Q_t$ is the per capita sentence length ("corrections") budget, and X_t is the exit rate of a prisoner selected at random in period t (cf., Raphael and Stoll 2009).[24] The prices w_t and r_t are taken to be exogenous. I will refer to equations (24), (25), and (26) as the crime equation, the budget equation, and the prison equation, respectively.

The crime equation clarifies that crime is mechanically reduced by imprisonment; this is the incapacitation effect of prison that is discussed in the literature. The budget equation clarifies that governments must pay for furnishing a probability of apprehension, regardless of whether the crime rate is high or low. However, governments do not have to pay for long prison sentences if crime falls by enough to keep prison populations low.[25] Intuitively, the most effective punishment is the one that never needs to be carried out.

[24]For the purposes of the theory described here, I consider judicial expenditures part of the corrections budget. It is useful to think of w_t as the price of an arrest and r_t as the price of incarcerating a person for a year. For example, if the typical police officer arrests 10 people a year and costs the government \$100,000 in wages, benefits, and so on, then $w_t = \$10,000$. A typical estimate of r_t is \$20,000 (Bureau of Justice Statistics 2004).

[25]This issue was treated formally in Blumstein and Nagin (1978).

This idea, immediately recognized by every parent, is quite old in the crime literature and dates at least to Bentham (1789).[26]

At first blush, it seems that if we impose parametric assumptions on the sentence length distribution so that $P(S_t \geq s)$ is a function of a finite number of parameters, then this problem can be solved analytically with standard recursive methods.[27] This is somewhat illusory, however. To understand the nature of the difficulty, observe that those in prison in a given period were either free, engaged in crime, caught, and sentenced to at least 1 period in prison as of 1 period ago, or were free, engaged in crime, caught, and sentenced to at least 2 periods in prison as of 2 periods ago, and so on. This is just the cohort distribution of those in prison. Formally,

$$Q_t = \sum_{s=1}^{\infty}(1 - Q_{t-s})G_{t-s}p_{t-s}P(S_{t-s} \geq s) \tag{27}$$

This leads to

$$\begin{aligned} Q_t &= (1 - Q_{t-1})G_{t-1}p_{t-1} + \sum_{s=1}^{\infty}(1 - Q_{t-1-s})G_{t-1-s}p_{t-1-s}P(S_{t-1-s} \geq s+1) \\ &\equiv (1 - Q_{t-1})G_{t-1}p_{t-1} + Q_{t-1}(1 - X_{t-1}) \end{aligned} \tag{28}$$

with

$$X_t = \frac{\sum_{s=1}^{\infty}\omega_{t-s}h_{t-s}}{\sum_{s=1}^{\infty}\omega_{t-s}} \tag{29}$$

[26] A more recent discussion, with many interesting examples, is given by Kleiman (2009).
[27] For an introduction to recursive methods, see Adda and Cooper (2003).

where $h_{t-s} = P(S_{t-s} = s)/P(S_{t-s} \geq s)$ is the hazard of exiting prison at period t for a prisoner from entry cohort $t - s$. This formulation emphasizes that the exit probability depends on all prior sentencing choices made by government. It is difficult to see how this problem could be made recursive with a finite state vector.

Adopting a normative perspective that the government should care equally about each generation, however, it is possible to characterize the government's solution cleanly. When the government cares about each generation equally, β is arbitrarily close to 1, and the objective $E_0 \left[\sum_{t=0}^{\infty} \beta^t C_t\right]$ is proportional to steady state crime.[28]

To begin, I define the sense in which the term "steady state" is used.

A1 (First Moment): For every t, $E[S_t]$ exists.

A2 (Steady State): For every t, $p_t \equiv p$, $S_t \equiv S$, $b_t^* \equiv b^*$

Under A1 and A2, we have the following results.

Result 1: $Q_t \equiv Q = GpE[S]/(1 + GpE[S])$, where $E[S]$ is the expected sentence length.

Proof: First, note that $E[S]$ can be rewritten as $\sum_{s=1}^{\infty} P(S \geq s)$. This can be seen by rearranging the terms of the implicit triangular sum. Rearrangement does not affect the limit since all terms are non-negative. Then note that in steady state, $Q = (1 - Q)Gp \sum_{s=1}^{\infty} P(S \geq s) = (1 - Q)GpE[S]$ by equation (27). Rearrangement then yields the result. □

[28] This point is also made in Manning (2003, Chapter 2).

Result 2: $C_t \equiv C = G/(1 + GpE[S])$, where C_t denotes the number of crimes per person. This implies that $Q = CpE[S]$: the fraction of people in prison equals the probability of crime, times the probability of arrest, times the expected number of periods detained.

Proof: By definition, $C = (1 - Q)G$. Then $1 - Q = 1/(1 + GpE[S])$ by Result 1. □

An important use of Results 1 and 2 is a decomposition of crime elasticities with respect to p and $E[S]$ into deterrence and incapacitation components.

Result 3: In steady state, (i) the elasticity of crime with respect to the probability of apprehension can be decomposed into deterrence and incapacitation components; (ii) the elasticity of crime with respect to the expected sentence length can be likewise decomposed; and (iii) the incapacitation components are equal. Formally,

$$\varepsilon_p = (1 - Q)\eta_p - Q \tag{30}$$

$$\varepsilon_{E[S]} = (1 - Q)\eta_{E[S]} - Q \tag{31}$$

Proof: Follows from Results 1 and 2, calculus, and algebra. □

Result 3 means that the overall crime elasticities, $\varepsilon_{E[S]}$ and ε_p, are both weighted averages of the deterrence elasticities, $\eta_{E[S]}$ and η_p, and an incapacitation elasticity of -1. The weights in the decomposition are $1 - Q$ and

Q. Moreover, Result 3 indicates that incapacitation effects are equal for p and $E[S]$. Thus, incapacitation is never a good argument for allocating the marginal criminal justice dollar to funding sentence enhancements instead of improvements to the probability of apprehension—the same incapacitation benefit could be generated by spending more money on increasing the probability of apprehension, such as by increasing funding for police.[29] Intuitively, prisoners must first be apprehended before they can be incapacitated by prison.

With this characterization of steady state crime, we can solve the government's problem when the government appropriate cares equally about all generations and hence minimizes steady state crime. Note that in steady state, government cannot run a deficit and simply spends a constant amount each period. These considerations imply that when the government cares equally about all generations, the dynamic problem described in (23) can be simplified to

$$\min_{p,E[S]} C \quad \text{s.t.} \quad B \leq B_0 \tag{32}$$

where $B = wp + rQ = wp + rCpE[S]$ is steady state expenditures, C is steady state crime, and B_0 is the maximum budget.

This problem is highly similar to a standard consumer or producer theory problem, but involves a nonlinear budget equation. A familiar conclusion is

[29] Additional mechanisms for increasing p through government investments include parole officers, information technology, state and federal crime labs, and so on.

that any interior solution to such a problem can be characterized as

$$\frac{\frac{\partial C}{\partial p}\frac{p}{C}}{\frac{\partial C}{\partial E[S]}\frac{E[S]}{C}} = \frac{\varepsilon_p}{\varepsilon_{E[S]}} = \frac{\frac{\partial B}{\partial p}\frac{p}{B}}{\frac{\partial B}{\partial E[S]}\frac{E[S]}{B}} \tag{33}$$

In words, the crime reducing benefit of spending 1 percent more on expected sentence lengths must be equal to the crime reducing benefit of spending 1 percent more on the probability of apprehension. This is the elasticity form of the classic conclusion that the marginal benefit of increased spending on any given budget item must equal the marginal benefit of increased spending on any other budget item.

In addition to the interior solutions to this problem, there are also corner solutions. Abstract from the corner solutions analogous to quasilinear preferences in consumer theory.[30] There are also more interesting corner solutions that are associated with the nonlinearity of the government budget constraint. These nonlinearities are due to the fact that the crime rate, C, enters the government budget constraint directly, due to the influence of the crime rate on the prison population (Blumstein and Nagin 1978).

To develop this idea, observe that the percentage increase in expenditures,

[30] A formal characterization of when it is reasonable to rule out corner solutions of this type would take me far afield. Any formal model of behavior, such as the model of offender behavior outlined above, can be used to provide such a characterization.

given a percentage increase in p and $E[S]$, is

$$\frac{\partial B}{p}\frac{p}{B} = 1 + (1-\sigma)\varepsilon_p \tag{34}$$

$$\frac{\partial B}{\partial E[S]}\frac{E[S]}{B} = (1-\sigma)(1+\varepsilon_{E[S]}) \tag{35}$$

where $\sigma = \frac{wp}{B}$ is the policing share of the criminal justice budget, and note that both of these quantities can be negative—that is, for both p and $E[S]$, the investments can be self financing.

This result stands in contrast with the standard linear budget, where both budget elasticities are equal to 1. Here, the budget elasticities are smaller than 1, because of the possibility that crime is reduced by the government's investments in crime control. If crime is reduced enough, then both budget elasticities are negative. Thus, under a maintained assumption of cost minimization on the part of government, we should not expect to observe crime elasticities of large magnitude. If they were to exist, arguendo, then government would have recognized that crime could be lowered and money could be saved, by investing more. On the other hand, when cost minimization is not a maintained assumption, this kind of reasoning suggests an avenue for substantial government savings. Increasing the probability of apprehension will lower crime and save money if $\varepsilon_p < \frac{-1}{1-\sigma}$. Increasing expected sentence lengths will lower crime and save money if $\varepsilon_{E[S]} < -1$.

There is also a further corner solution that arises due to the nonlinear budget. This occurs when offenders are more responsive to a 1 percent increase

in $E[S]$ than they are to a 1 percent increase in p, or when $|\varepsilon_{E[S]}| > |\varepsilon_p|$. However, it is hard to imagine that behavior is more elastic with respect to sentences than the probability of apprehension, regardless of p and $E[S]$. More plausibly, behavior is more elastic with respect to the probability of apprehension, and increasingly so when sentences become long. For example, plausible calibrations of equations (11) and (12) suggest strong the probability of apprehension and sentence lengths and thus the implausibility of this type of corner solution.

Returning to the characterization of an interior solution, we see that at an interior optimum, the fraction of the criminal justice budget that should be devoted to policing is given by

$$\sigma = 1 - \frac{\varepsilon_{E[S]}}{\varepsilon_p} \tag{36}$$

Armed with estimates of deterrence elasticities, this equation gives a simple formula for optimal allocation of government resources.

FIGURE 2 ABOUT HERE

Figure 2 shows the time series of federal, state, and local criminal justice expenditures devoted to non-policing sources (i.e., judicial expenses are here viewed as part of detention rather than apprehension). The figure shows plainly the increased criminal justice focus on detention rather than apprehension over the past 40 years, particularly in the period 1970-1995.

Equation (36) shows that, maintaining the hypothesis that the federalist

system achieves cost minimization, policing must have become less effective over this period. To understand the magnitude of the implied effect, suppose that arrests per officer are constant. Then in 1970, equation (36) shows that the hypothesis of cost minimization implies an elasticity of crime with respect to police, relative to that with respect to sentence lengths, of about 2.5. If cost minimization were true, then it must be the case that over the period 1970-2005, the elasticity of policing, relative to that of sentence lengths, fell from 2.5 to about 1.8, or a decline of about one-third. To the extent that such a decline seems implausible, it suggests that either money could *currently* be saved by reallocating spending away from prisons and towards police, or that money could *historically* have been saved by reallocating spending away from police and towards prison.

If potential criminals are completely unresponsive to punishment parameters, so that ε_p and $\varepsilon_{E[S]}$ are both zero, then the government's problem as stated is ill-posed: it then always makes sense to maximize the size of the prison population, and for a fixed budget, this is always most efficiently done by cutting police and increasing $E[S]$. The government's problem would also be ill-posed if, in the current environment, the elasticity of crime with respect to sentence lengths were elastic, i.e., if $\eta_{E[S]} < -1$.[31] Under such a scenario, equation (35) shows that increasing sentences is self-funding. The hypothesis of cost minimization thus implies an inelastic response of crime to expected sentence lengths.

[31] Equivalently, if $\varepsilon_{E[S]} < -1$.

The steady state calculations above are helpful in understanding the incentives facing governments, but miss a key dynamic insight. Suppose government is facing a tight budget, but is trying to get crime under control. As emphasized by equation (25), increasing police in period t generates deterrence in period t, but also costs money in period t. In contrast, increasing sentence lengths in period t generates deterrence in period t, but costs are not borne until period $t+1$ and do not become large until Q_t becomes large. This will frequently be many periods later. As discussed in Klick and Tabarrok (2005), for example, the three-strikes law passed in California in 1994 sentenced a large number of offenders to 25 years to life instead of the typical sentence of 10 years. Passing three-strikes thus postponed paying for crime control for 10 years. This is presumably nearly always an attractive option for a state government facing a balanced budget requirement. Indeed, from this perspective, and in light of revenue cycles, balanced budget requirements presumably distort the criminal justice policy choices of any government with a bias in favor of the current generation.[32]

5 Conclusion

Many interesting features of crime and crime control are dynamic in nature. In recent years, a small literature has emerged that addresses some of these

[32] While I earlier adopted a normative perspective that government should care equally about each generation, this is somewhat at odds with what one would expect from term limits and electoral uncertainty, both of which would presumably lead politicians to behave as if they had a taste for the present.

issues. Key areas of focus in the existing literature include (1) the difficulty of controlling, via changes to sentencing, the criminal involvement of those with short time horizons; (2) the intertemporal substitution of criminal activity; (3) the accumulation of human capital, both criminal and legitimate, that can lead to important differences between short- and long-run crime reduction benefits of policy interventions.

An important dynamic feature of crime that deserves more attention is the government's problem of how best to allocate criminal justice expenditures over time and between uses. This issue has acquired a renewed relevance this past year with many governments seeking to cut costs in the wake of the financial crisis. A particularly compelling question is the optimal division of criminal justice dollars between police and prisons. I have emphasized that the government's solution to this problem relates to time preferences for marginal offenders, which powerfully influence the relative elasticities of crime with respect to the probability of apprehension and with respect to an expected sentence length. The United States has substantially reduced the fraction of criminal justice dollars devoted to policing over the last 40 years and devoted the marginal dollar instead to corrections expenses. If the marginal offender has short time horizons, then this may well not be the most efficient use of the marginal criminal justice dollar, suggesting the possibility of substantial government savings.

Finally, I suspect that an important future research direction is inducing criminals to provide information to government that is useful for crime con-

trol. Currently, punishments depend primarily on criminal history and, to a lesser extent, judicial discretion. An interesting question is whether government can elicit offender beliefs about the likelihood of recidivism. Assuming those beliefs are accurate, such information could be highly valuable. For example, several states are currently considering engaging in release of prisoners. A natural question is who should be released. The obvious answer is those with low probabilities of recidivism—but it is hard to know who those prisoners are. One example of a policy that could be used to elicit beliefs about recidivism probabilities is as follows. Those with 1 year left on their sentence are offered early release, contingent upon being willing to wear an ankle bracelet. An offender anticipating recidivating might be loathe to agree to such terms, because facing a higher probability of apprehension could lead to a longer prison term than 1 year.[33] Economists have not yet begun to carefully think about what kinds of information revelation policies would be useful for government crime control, but we should be at the forefront of such efforts.

[33]Of course, offenders could incorrectly anticipate abstaining from crime. Offender misperceptions of future criminal propensity would limit the value of such information revelation policies.

References

Adda, J., and R. W. Cooper (2003): *Dynamic Economics: Quantitative Methods and Applications*. MIT Press, Cambridge.

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

Bentham, J. (1789): *An Introduction to the Principles of Morals and Legislation*. Clarendon Press, Oxford.

Blinder, A. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8(4), 436–455.

Blumstein, A. (2002): "Crime Modeling," *Operations Research*, 50(1), 16–24.

Blumstein, A., and D. Nagin (1978): "On the Optimum Use of Incarceration for Crime Control," *Operations Research*, 26(3), 381–405.

Braga, A. (2005): "Hot Spots Policing and Crime Prevention: A Systematic Review of Randomized Controlled Trials," *Journal of Experimental Criminology*, 1(3).

Burdett, K., R. Lagos, and R. Wright (2003): "Crime, Inequality, and Unemployment," *American Economic Review*, 93(5), 1764–1777.

——— (2004): "An On-the-Job Search Model of Crime, Inequality, and Unemployment," *International Economic Review*, 45(3), 681–706.

Bureau of Justice Statistics (2004): "State Prison Expenditures, 2001," *Bureau of Justice Statistics Special Report*.

Cohen, J. (1986): "Appendix B. Research on Criminal Careers: Individual Frequency Rates and Offense Seriousness," in *Criminal Careers and "Career Criminals"*, ed. by A. Blumstein, J. Cohen, J. A. Roth, and C. A. Visher, vol. 1, pp. 292–418. National Academy Press, Washington, D.C.

Devroye, L. (1986): *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

Di Tella, R., and E. Schargrodsky (2004): "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack," *American Economic Review*, 94(1), 115–133.

DiNardo, J. E., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64(5), 1001–1044.

Eiserer, T. (2005): "Police to Create Task Force to Combat Crime Hot Spots Effort to Start in Early July with About 60 Officers," *Dallas Morning News*.

Flinn, C. (1986): "Dynamic Models of Criminal Careers," in *Criminal Careers and "Career Criminals"*, ed. by A. Blumstein, J. Cohen, J. A. Roth, and C. A. Visher, vol. 2, pp. 356–379. National Academy Press, Washington, D.C.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996): "Crime and Social Interactions," *Quarterly Journal of Economics*, 111(2), 507–548.

Grogger, J. (1998): "Market Wages and Youth Crime," *Journal of Labor Economics*, 16(4), 756–791.

Gronau, R. (1977): "Leisure, Home Production, and Work: The Theory of the Allocation of Time Revisited," *Journal of Political Economy*, 85(6), 1099–1123.

Heinzmann, D. (2004): "Murders Drop, but Chicago Leads U.S. ; Rate Down Sharply in Last 6 Months," *Chicago Tribune*.

Higgins, J. (2006): "City Police to Step up Presence: Mayor Adds $200,000 in Overtime to Target Hot Spots, Enforce Curfew," *Akron Beacon Journal*.

Hoover, M. (2007): "Sweeping Away Crime; Cooling Down Hot Spots for Blight; Sweeping Away the Crime; City Cops Try to Rid Area of Rroubles, but Residents Don't Think It's Enough," *York (PA) Daily Record*.

Huang, C.-C., D. Laing, and P. Wang (2004): "Crime and Poverty: A Search-Theoretic Approach," *International Economic Review*, 45(3), 909–938.

Hunt, A. (2009): "Task Force Aims to Cool Down Crime Hot Spots in Detroit," *Detroit Free Press*.

Imai, S., and K. Krishna (2004): "Employment, Deterrence, and Crime in a Dynamic Model," *International Economic Review*, 45(3), 845–872.

Imrohoroglu, A., A. Merlo, and P. Rupert (2004): "What Accounts for the Decline in Crime?," *International Economic Review*, 45(3), 707–729.

Jacob, B., L. Lefgren, and E. Moretti (2007): "The Dynamics of Criminal Behavior: Evidence from Weather Shocks," *Journal of Human Resources*, 42(3), 489–527.

Judd, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge.

Katz, M. (2009): "Camden's New Policing Style is Both Loved, Hated," *Philadelphia Inquirer*.

Kleiman, M. (2009): *When Brute Force Fails: Strategy for Crime Control*. Princeton University Press, Princeton.

Klick, J., and A. Tabarrok (2005): "Using Terror Alert Levels to Estimate the Effect of Police on Crime," *Journal of Law and Economics*, 48.

Lawless, J. F. (2003): *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, Hoboken.

Lee, D. S., and J. McCrary (2005): "Crime, Punishment, and Myopia," *NBER Working Paper # 11491*.

——— (2009): "The Deterrence Effect of Prison: Dynamic Theory and Evidence," Unpublished manuscript, UC Berkeley.

Lemieux, T., B. Fortin, and P. Frechette (1994): "The Effect of Taxes on Labor Supply in the Underground Economy," *American Economic Review*, 84(1), 231–254.

Ljungqvist, L., and T. J. Sargent (2004): *Recursive Macroeconomic Theory*. MIT Press, Cambridge.

Lochner, L. (2004): "Education, Work, and Crime: A Human Capital Approach," *International Economic Review*, 45(3), 811–843.

Manning, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press, Princeton.

McCall, J. J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84(1), 113–126.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.

Piehl, A. M., and J. J. DiIulio, Jr. (1995): "'Does Prison Pay?' Revisited: Returning to the Crime Scene," *Brookings Review*, pp. 21–25.

Raghavan, S. (2005): "Johnson Targets Crime Hot Spots," *Washington Post*.

Raphael, S., and M. A. Stoll (2009): "Why Are So Many Americans in Prison?," in *Do Prisons Make Us Safer?: The Benefits and Costs of the Prison Boom*, ed. by S. Raphael, and M. A. Stoll, pp. 27–72. Russell Sage Foundation, New York.

Rust, J. (1996): "Numerical Dynamic Programming in Economics," in *The Handbook of Computational Economics*, ed. by H. M. Amman, D. A. Kendrick, and J. Rust, vol. 1, pp. 619–729. Elsevier, Amsterdam.

Sickles, R. C., and J. Williams (2008): "Turning from Crime: A Dynamic Perspective," *Journal of Econometrics*, 145(1-2), 158–173.

Strotz, R. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23(3), 165–180.

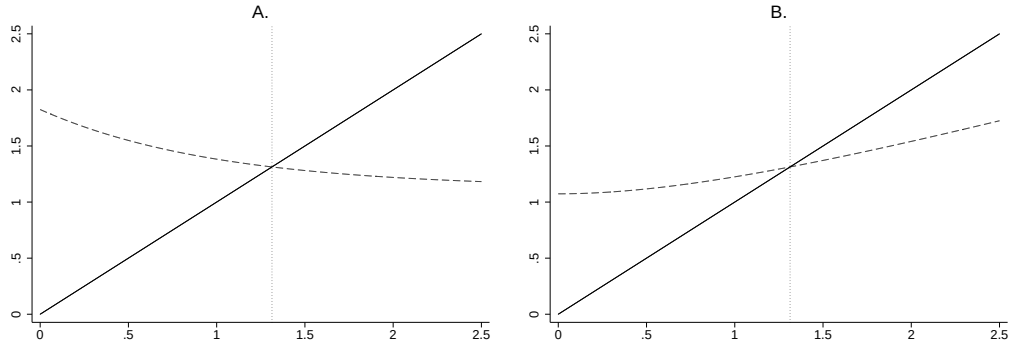
Viscusi, W. K. (1986): "The Risks and Rewards of Criminal Activity: A Comprehensive Test of Criminal Deterrence," *Journal of Labor Economics*, 4(3), 317–340.

Visher, C. A. (1986): "The Rand Inmate Survey: A Reanalysis," in *Criminal Careers and "Career Criminals"*, ed. by A. Blumstein, J. Cohen, J. A. Roth, and C. A. Visher, vol. 2, pp. 161–211. National Academy Press, Washington, D.C.

Williams, J., and R. C. Sickles (2002): "An Analysis of the Crime as Work Model: Evidence from the 1958 Philadelphia Birth Cohort Study," *Journal of Human Resources*, 37(3), 479–509.

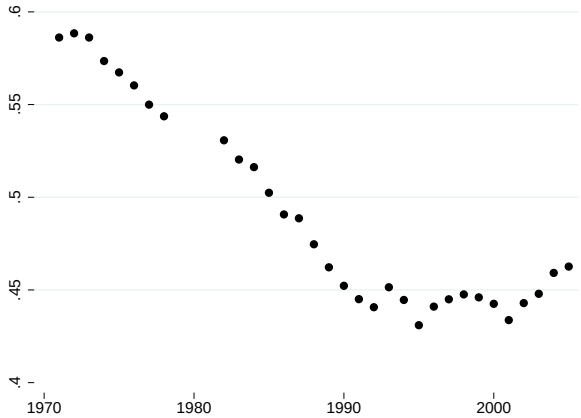
Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge.

Figure 1. Reservation Benefit as Fixed Point:
A. $b^* = \alpha + \beta\psi(b^*)$ and B. $b^* = \tilde{\alpha} + \tilde{\beta}\tilde{\psi}(b^*)$



Note: Figures drawn using baseline model with $a = 0$, $c = 1$, $\delta = 0.95$, $p = 0.2$, $\pi_s = (1-\gamma)\gamma^{s-1}$, with $\gamma = e^{-1/5}$, and $F(b) = 1 - \exp(-b)$.

Figure 2. Policing Budget Share, or σ: 1971-2005



Source: Bureau of Justice Statistics, Expenditure and Employment Series