

NBER WORKING PAPER SERIES

CRIME AND SOCIAL INTERACTIONS

Edward L. Glaeser
Bruce Sacerdote
José A. Scheinkman

Working Paper No. 5026

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1995

The National Science Foundation provided generous financial support. Helpful comments were provided by participants in the Latin American Econometric Society meetings, the University of Chicago Applications Workshop and the Hoover Institution economics workshop. Glenn Ellison, H. Ichimura, Guido Imbens, Richard Posner and Jesus Santos provided helpful suggestions. Julie Khaslavsky, Andrei Scheinkman and Jake Vigdor provided excellent research assistance. This paper is part of NBER's research programs in Growth and Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1995 by Edward L. Glaeser, Bruce Sacerdote and José A. Scheinkman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

CRIME AND SOCIAL INTERACTIONS

ABSTRACT

The high degree of variance of crime rates across space (and across time) is one of the oldest puzzles in the social sciences (see Quetelet (1835)). Our empirical work strongly suggests that this variance is not the result of observed or unobserved geographic attributes. This paper presents a model where social interactions create enough covariance across individuals to explain the high cross-city variance of crime rates. This model provides a natural index of social interactions which can compare the degree of social interaction across crimes, across geographic units and across time. Our index gives similar results for different data samples and suggests that the amount of social interactions are highest in petty crimes (such as larceny and auto theft), moderate in more serious crimes (assault, burglary and robbery) and almost negligible in murder and rape. The index of social interactions is also applied to non-criminal choices and we find that there is substantial interaction in schooling choice.

Edward L. Glaeser
Department of Economics
113 Littauer Center
Harvard University
Cambridge, MA 02138
and NBER

Bruce Sacerdote
Department of Economics
Harvard University
Cambridge, MA 02138

José A. Scheinkman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637

Quelquefois aussi le crime prend sa source dans l'esprit d'imitation, que l'homme possède á un haut degré et qu'il manifeste en toutes choses. A. Quetelet, *Du Système Social*

I. Introduction

There are seven serious crimes per one hundred citizens per year in the United States; this figure is more than double the serious crime rate of 1970.¹ 23% of African-American males aged 20-29 are currently under correctional supervision (i.e. in jail, on parole or on probation). One, admittedly alarmist, estimate of the total damage from crime in the U.S. is \$425 billion.² In several 1993 surveys, crime was listed as America's largest social problem.³ The study of crime has obvious policy relevance. Moreover, since criminal activity is so omnipresent and violent crime has such a vast history, the study of crime is intrinsically important to social science.

The modern economics literature on crime essentially follows from Becker's (1968)'s model of rational criminal activity. In this model, crime results when the benefits of crime to a criminal outweigh the cost of potential punishment. A long empirical literature (e.g. Ehrlich (1975), Levitt (1994)) followed testing of one of the comparative statics that falls naturally out of Becker's model: higher levels of deterrence produce lower quantities of crime.⁴ Much of the economics literature's focus on the relevance of deterrence seems to be a rebuttal of the even larger sociological and psychological literatures on "deviant behavior" which often claim that rationality and deterrence do little to explain levels of crime.⁵

¹FBI uniform crime reports shows a per capita crime rate of .067 in 1985 and .0773 in 1986. These can be compared with a per capita crime of .032 in 1970. We will use the uniform crime reports, not victimization studies, throughout this paper, due to lack of data availability of victimization studies with sufficient geographical information. Victimization studies are generally seen as more accurate, and changes in reporting practices, may account for some of the intertemporal variation in crime rates. However, there does not seem to any evidence to suggest cross sectional variation declines substantially when victimization studies are used.

²This estimate is from *Business Week*, December 13, 1993.

³The *Chicago Tribune*, September 10, 1994, reported that in a mid-July, 1994, New York Times/CBS News Poll, crime was listed as the nation's most important problem. Further survey results are described in *U.S. News and World Report*, February 7, 1994.

⁴There is also an extensive literature on crime's relation to economic opportunity, e.g. Freeman (1991).

⁵Wilson and Herrnstein (1980) provide a valuable introduction to these literatures.

We believe that the most puzzling aspect of crime is not its overall level, or that level's relationship (or non-relationship) with the observed quantities of deterrence.⁶ Rather, following Quetelet (1835), we believe that the most inexplicable aspect of crime is its large variance across time and space. The media trumpets the astounding rise in crime since the 1960s, but there are also cases where criminal activity seems to have fallen dramatically over time. From 1933 to 1961, homicide rates fell in half in the United States.⁷ Lane (1979) documents an equally substantial drop in homicides in Philadelphia over the late 19th century.

As large as the intertemporal differences in crime rates are, these differences are dwarfed by the differences in crime across space. There is a wide spectrum of average homicide rates across nations ranging from 11.7 homicides per million in Sweden to 105.3 homicides per million in the United States in 1980.⁸ Within the United States cities range widely in their crime rates from .008 serious crimes per capita for Ridgewood Village, New Jersey, to .384 serious crimes per capita for nearby Atlantic City.⁹ Even within a single city, the diversity across sub-city units can be astounding; the 123rd precinct of New York City has 2.16 crimes per hundred citizens while the 1st district had 20.59 crimes per hundred citizens.

If agents' decisions are independent, then city crime levels represent averages of large numbers of independent decisions. Elementary statistics tells us that these averages should be free of the effects of random idiosyncratic error terms and they should be close to the expected population mean. The great differences in crime rates across space must be explained by sufficiently different economic conditions (or levels of deterrence) in the different areas.

⁶In part these topics are less puzzling to us because of the extensive work that has been done on them.

⁷ From pre-1960 data, or for international comparisons, homicides provide the only reliably uniform definition of crime. Archer and Gartner (1984) provide an excellent discussion of cross-national crime studies.

⁸Because of the recession, 1980 was a particular bad year for U.S. homicide rates. Typical rates are between 98-100 homicides per million per year.

⁹As we will discuss later, Atlantic City's crimes per capita is particular high because its population does not include much of the large tourist population that enters the city but that is not counted as part of the city's population. The highest crime levels in a non-tourist city is Portland, Oregon, with .169 serious crimes per capita.

Indeed poor, inner city precincts generally have more crime than precincts in middle class Staten Island. However, even casual empiricism suggests that differences in observable local area characteristics can account for little of the variation in crime rates across cities in the U.S. or across precincts in New York city.

East Point, Georgia has a crime rate of 9.2 crimes per hundred. El Dorado, Arkansas, has a crime rate of 3.9 crimes per hundred. East Point has lower unemployment, a better educated populace, less poverty and higher income per capita. The 51st precinct of New York City has 4.6 crimes per hundred citizens while the wealthier 49th district has 11.6 crimes per capita. How can the radical differences in the crime rates of these areas be accounted for by their underlying economies? More rigorously, we generally find that less than 30% of the variation in cross-city or cross-precinct crime rates can be explained by differences in local area attributes.¹⁰

The basic requirement for justifying a high variance of crime rates across space (over and above the variance predicted by cross-sectional differences in local conditions) is a positive covariance across agents' decisions about crime. When one agent's decision to become a criminal positively affects his neighbor's decision to enter a life of crime, then there is no reason why an aggregate figure (such as urban crime levels) should be close to the population mean of that figure. Positive interactions across agents' decisions to become criminals predict that cities' crime rates will differ from the rates predicted by the cities' basic characteristics, and that crime rates will differ substantially across locations. Our empirical results support the existence of these positive interactions and comfortably reject the possibility that decisions about crime are independent within a local area.¹¹

¹⁰The regressions that produce these results will be discussed later in the paper. In general we find that controlling for local area attributes brings us an R^2 of between 25-30% for New York City precincts and slightly over 30% for cities across the U.S. in 1985. Cross-city regressions in 1970 yield somewhat higher R^2 's, but still the remaining variance is far too high to be compatible with a model where agent's decisions about crime are independent.

¹¹Our work on this area is supportive of the findings of Case and Katz (1992) who find local interactions using a survey of Boston youth.

In order to make sense of the covariance across agents that we find in the data, we present two models that build on the previous work on social interactions and crime (e.g. Sah (1990), and Murphy, Shleifer and Vishny (1993)). Previous work has shown that positive interactions can produce high variances of crime rates across cities, and has presented candidate mechanisms for explaining these interactions. However, most of this work tends to predict a finite number (usually two) of distinct distributions of crime rates over space (which does not seem to be supported in the data).¹² This work also does not present a measure of the degree of social interactions that can be used in either hypothesis testing or in cross-crime, cross-space and cross-time comparisons.¹³ Our models predict distributions that are similar to those in the data.¹⁴ Moreover, our models give us estimable parameters that provide us with a natural index of the degree of social interaction.

This paper presents two models where agents are arranged on a lattice and agents' decisions to participate in crime are a function of their own attributes and of their neighbors' decisions about criminal activities.¹⁵ In these models, the population is divided into two classes of agents: (1) agents who influence and are influenced by their neighbors and (2) agents who influence their neighbors, but who are not themselves influenceable ("fixed agents"). When we examine the limiting distributions of crime rates that are determined by these models, we find that the variance across localities will be a function of the number of fixed agents.

Fixed agents lower the degree of interactions across agents and hence decrease the expected variance of crime rates across cities. In the extreme this is obvious, as all agents become fixed, criminal decisions lose their interdependence and crime rates become the average of independent

¹²The one major example of a prior model of multiple equilibria that is empirically fit in the crime area is Smith and Wright (1993) which does indeed find two distinct distributions in their data.

¹³These comments are not criticisms of the earlier models which were not motivated as empirical models, but rather to illustrate how social interactions can be generated and how they can effect the variance of crime rates over space.

¹⁴In part our ability to fit the data better comes about because we have an extra degree of freedom, the percentage of fixed agents, in our model.

¹⁵These model are derived from the voter models (see Kindermann and Snell (1980)). Many of our results are related to the discussion in Scheinkman and Woodford (1993).

decisions.¹⁶ Both models provide the same critical result: the variance of crime rates (times the square root of population) across a sample of cities is equal to a nonlinear function of the proportion of fixed agents in each city time the variance of crime rates (weighted by the square root of population) that one would expect if all decisions were independent times.¹⁷

We use this fact empirically and estimate for a variety of crimes in the U.S. the proportion of fixed agents predicted by the variance of crime rates across cities. This proportion of fixed agents provides us with an index of the degree of social interaction, and we can ask how the level of social interaction changes across crimes or over time. The number of fixed agents can be interpreted in many ways: (1) the expected distance between two fixed agents is the expected size of a group with positive social interactions, so the number of fixed agents determines the average social group size, (2) fixed agents can be viewed as agents who do not observe their neighbors actions, so the number of fixed agents may reflect the share of the population that is not connected to their neighbors, and (3) the number of fixed agents can just be seen as a metaphor for the forces that slow social interaction. Forces that slow social interaction among potential criminals can include strong parents, or formal schooling, or any force that provides information that counters criminal influences.

The empirical section of the study presents this index of interactions for a variety of different crimes in the U.S. in 1985, in 1970 and across New York City in 1985. In all three samples, we find a high degree of interaction for larceny and auto theft. Our data shows moderate (but still large) levels of interaction for assault, burglary and robbery. There are very low levels of social interaction for arson, murder and rape. While there is some difference in these measures over time (the level of social interaction seems to have risen between 1970 and 1985), overall we find that the levels of interaction are fairly similar across the three samples. We believe that this similarity (which

¹⁶In fact, fixed agents determine the variance of crime rates not so much because they are uninfluenced by their neighbors, but rather because they break the chains of influence within the locality. Fixed agents serve as "circuit breakers" in the transmission of criminal tendencies.

¹⁷ We assume, as least until Table 5, that there is the same number of fixed agents in each city.

does not hold for the mean levels of these crimes) supports the usefulness and reliability of our methodology.

In a further test of our methodology, we apply it to the percent of a city which was female; that percentage should presumably not be a decision variable determined by social interactions. With a trivial correction for the heterogeneity in the cities' attractiveness to the different sexes, we found no social interactions in this variable. We also used our methodology to examine schooling, which displayed substantial quantities of social interactions. Our methodology, in principle, can be used to test for cross-individual effects in many variables beyond crime.

We also present a variety of improvements in the basic methodology. We allow for repeat offenders. We estimate the amount of cross-city variance that can be explained by variance in the underlying characteristics of cities and give lower bounds on the amount of social interaction that must be occurring. We allow the proportion of fixed agents (the degree of interaction) to fluctuate with city characteristics. We find that for serious crimes in general, for larceny and for auto theft, the degree of social interactions is larger in those communities where families are less intact. Finally, we show that several of the usual interaction mechanisms are only dimly supported by the data. The mechanism of congestion in law enforcement is particularly unsupported by cross-city or cross-precinct evidence.

2. Previous Literature and Positive Interactions

A number of models already explain why seemingly identical cities can have different levels of crime (or rent-seeking more generally). Sah (1990) represents a benchmark version of the existing models explaining the puzzle of high cross-sectional variance in crime. In this work, a large number of criminals congest the law enforcement mechanism. Police cannot arrest more than a fixed number of criminals, so when there is too much crime, the probability of being arrested goes down. As this happens, the cost of being a criminal also falls. Two (or more) equilibria can result: one equilibria with high crime levels and low probabilities of arrest and the other equilibria with low crime levels and a high probability of arrest.

Several other versions of this basic model also exist. Murphy, Shleifer and Vishny (1992) suggest an alternative mechanism where high levels of criminal (or rent-seeking) behavior crowd out legal activities. As the number of criminals rise, the returns from not being a criminal falls because legal revenues are stolen by criminals. Again, two equilibria can result. In one equilibria, agents follow legal activities primarily and the returns from legal activities are high. In the second equilibria, agents are primarily rent-seekers taking from the legal workers, thus making the returns from legal activities low and inducing agents to prefer crime.

Alternative models can generate multiple equilibria if the non-criminals are involved in disciplining the criminals,¹⁸ or in deciding (with their votes) the amount spent on crime prevention. As the number of non-criminals falls, the amount of resources allocated to preventing crime also falls and it becomes more advantageous to follow a life of crime. Multiple equilibria also can result if crime is stopped because crime is stigmatizing; as the number of criminals rise, the average criminal becomes a "normal" member of society. As crime rises, the stigma from crime falls and it becomes more attractive to commit crimes.¹⁹ Stigma may also work if higher levels of crime stigmatize an area and make outside employers less likely to hire the residents of a particular city; that lack of hiring then lowers the cost and increases the quantity of crime in the area.²⁰

While the multiple equilibria models do generate more variance across locations than do models with totally independent decisions, these models have some severe problems as empirical tools. Taken literally, the models predict that there will be two stable equilibria (or possibly some other small, finite number of equilibria) and urban crime rates will be grouped around two primary points. Empirically, these models predict that the data should be

¹⁸Jacobs (1961) is a primary advocate of the critical role of civilian policing in limiting urban crime.

¹⁹Glaeser (1992) presents multiple equilibria models based on stigma.

²⁰There are also many possible negative interactions, most classically standard competition for a scarce resource (victims) may mean that more criminals lowers the returns to criminal activity. However, these negative interactions would lower, not raise, the observed variance of crime rates.

clustered into two primary distributions, and that once we allow the data to come from two distributions, we should find little excess variance. A quick inspection of crime rates in Figures 2, 3 and 4 shows that the distribution of crime rates certainly do not suggest that cities to end up in one of a few possible equilibria.²¹ Also, as we will show in Table 6, there is still a great deal of excess variance once we have allowed the crime rates to come from two (or more) distributions.

In order to create more empirically palatable models, we will assume that the social interaction occurs at the very local level: each agent's decisions are influenced mainly by their neighbors' decisions. The multiple equilibria models have the complementarity work through the neighborhood average. By focusing on local rather than global interactions and by assuming that there exists a fraction of the population that does not respond to interactions, we will avoid the empirically unpalatable situation of predicting that cities' will converge to a few distinct equilibria. Instead, our models will simply predict a large variance of urban crime rates.

Rather than focus on the form of the interactions (as the multiple equilibria models do), we will assume a positive interaction across criminals and then focus on getting a more empirically plausible model. The current evidence (Case and Katz (1992), Crane (1990)) confirms a positive interaction across agents, but it is not conclusive about the form of the interaction. Most of the popular interactions (congestion in law enforcement, crowding out of legal activities) do not seem to fit the U.S. data well (as we will show in Table 7).

Of course, moving from a global interactions model to a local interactions model immediately suggests different justifications for the interaction. The traditional stories of interactions (congestion of law enforcement, elimination of legal opportunities) tend to be best suited for models of global interaction (although they could create local interactions as well). Justifications for local interactions must focus on agent-to-agent interactions. Among other things,

²¹In these figures, the crime rates (especially for 1970) look leptokurtotic and in fact the distributions of crime have non-Normal degrees of skewness and kurtosis. These features of the distributions occur to a large extent because of the relationship between crime rates and city size, and because city sizes are famously leptokurtotic (i.e. Zipf's law holds).

local interactions can plausibly come from: (1) information flows about the returns to crime or about how to be a criminal, (2) inputs from family members that determine the costs of crime or the taste for crime, (3) peer group approval or (4) the importance of monitoring by very proximate neighbors for deterrence (and those neighbors don't monitor when they are themselves criminals). We will remain slightly agnostic throughout the model and until Table 7 in the empirical work on the form of the positive interactions. Ultimately this paper dwells on measuring the presence of social interactions and their influence on crime, not on determining the exact form of social interactions.

3. Model

In this section we will present two similar but slightly different models. In the first model, agents' decisions will be determined randomly and local interactions will be modeled as a heightened propensity to engage in crime. In the second model, maximization will appear more plainly. Both models will suggest the same index of social interaction and the same interpretation of that index.

Model One -- Social Interactions and the Propensity to Commit Crimes

We present here a model that is capable of generating the high variability of crime incidence that we observe in the data. We assume that each agent is indexed by an integer i , and that each agent can choose one of two actions $\{1,0\}$. Action 1 will be interpreted as committing a crime. Although a dynamic model will be presented, we are really only interested in the properties of its stationary distribution. The dynamics are described only to allow us to compute properties of the stationary distribution.

At time 0, each agent chooses an action independently and the probability that an agent chooses action 1 is p . With probability $\pi > 0$, an agent i belongs to a set S of fixed agents. These agents will not change their action over time. The probability of being fixed is independent across agents.

The remaining agents will change their actions occasionally to conform to the choices made by their neighbors (this is the social interaction). The neighbors of an agent i are the agents in $N(i) = \{i-1, i+1\}$. Associated with each agent $i \notin S$, (i.e., all non-fixed agents) there is an independent Poisson process with mean time 1. At each epoch of the Poisson process associated with i , the action of the agent i changes to that of one of its neighbors with equal probability. This defines stochastic processes a_i^t , for each integer i . Shift to the neighbors' actions are not explicitly the result of maximizing behavior, but they can be seen as reflecting the gains from following one's neighbors. In the next model, we will show that this equivalent in its long run behavior to a model based on maximizing behavior.

For the given parameters (p, π) , there is a limit measure $\mu(p, \pi)$. More precisely, given an n , there exists a limit measure $\mu_n(p, \pi)$ defined over configurations of $\{a_i : |i| \leq n\}$, and if $m > n$, the measure induced over the configurations of $\{a_i : |i| \leq n\}$ by $\mu_m(p, \pi)$ agrees with $\mu_n(p, \pi)$. Further, the measure $\mu(p, \pi)$ is invariant.

To see this, first consider the measure $\mu(p, \pi)$ given the set S and the a_i^0 for each $i \in S$. Given any integer $i \notin S$, we write i_- for the largest element of S which is less than i , and i_+ for the smallest element of S which is greater than i . In other words, i_- and i_+ are the elements of S that "bracket" i . If $a_{i_-}^0 = a_{i_+}^0$ then as $t \rightarrow \infty$ it is easy to see that $a_i^t \rightarrow a_{i_-}^0$, with probability one. If an interval of non-fixed agents is bracketed by two fixed agents, both choosing action 1, then in the limit, all of the non-fixed agents in that interval will also choose action 1.

Hence $\mu(p, \pi)$ assigns probability one (conditional on $i_-, i_+ \in S$, and $a_{i_-}^0 = a_{i_+}^0$) to the set of configurations in which a_j is constant for $i_- \leq j \leq i_+$. If $a_{i_-}^0 \neq a_{i_+}^0$, then it is again easy to see that with probability approaching one, as $t \rightarrow \infty$, if $a_i^t = a_{i_-}^0$ then for any $i_- \leq j \leq i$, $a_j^t = a_{i_-}^0$. This result just means that on an interval of non-fixed agents between two fixed agents, the agents' states are "sorted," and only one agent (at a point in time -- this agent will not be fixed) will have a neighbor that takes an action different than his own. Everyone else (at a point in time) will have neighbors doing exactly what they are doing. Hence

$\mu(p,\pi)$ assigns probability one to the set of configurations that are sorted in this way.

One can use elementary Markov chain arguments to show further that if i_- and i_+ are consecutive points of S , with say $a_{i_-}^0 = 0$ and $a_{i_+}^0 = 1$, and $i_+ - i_- = l$ then the probability that $\mu(p,\pi)$ assigns (again, conditional on $i_-, i_+ \in S$, and $a_{i_-}^0 = 0$ and $a_{i_+}^0 = 1$) to the configurations in which actions 0 is chosen exactly by agents j , $i_- \leq j \leq i_+ - 1$, in the interval $[i_-, i_+]$, equals $1/l$.

As we mentioned above, we are really interested in the properties of the stationary distribution $\mu(p,\pi)$. We will therefore drop the time superscript t with the understanding that we are examining the properties of the stationary distribution. It is apparent from the dynamics that the average number of agents in a group that take a particular action does not change with the number of agents in the group. Hence, using the distribution $\mu(p,\pi)$, for any n , the expected number of agents $|i| \leq n$ that chooses action 1 is $p(2n+1)$. The presence of "fixed" agents guarantees that if $j < i$, as $|j-i| \rightarrow \infty$, then with a probability that tends exponentially to zero, the action chosen by j is independent of that chosen by i . Hence central limit behavior results,²² which means that:

$$\frac{1}{\sqrt{2n+1}} \sum_{|i| \leq n} (a_i - p) \rightarrow N[0, \sigma^2], \quad (3.1)$$

and further,

$$\sigma^2 = \text{var}(a_0) + 2 \sum_{i>0} \text{cov}(a_0, a_i) \quad (3.2)$$

We know that $\text{var}(a_0) = p(1-p)$. The computation of the covariance terms in (3.2) is elementary but lengthy, and we leave it for an appendix (available upon request). The important fact that we will use later is that

$$\sigma^2 = f(\pi)p(1-p) \quad (3.3)$$

²² See, for example, White (1984), Theorem 5.19, page 124.

where, $f(1)=1$, $f'(\pi)<1$ and, $\lim_{\pi \rightarrow 0} f(\pi)=\infty$.²³ In Figure 1, $f(\pi)$ is plotted next to the $f(\pi)$ that is generated by the next model. We also know that the variance of the appropriate normalized sum will converge to σ^2 at a speed of at least $1/n$.

The term $f(\pi)$ results from the presence of local interaction. If no imitators are present this term equals one and the variance is $p(1-p)$. As the number of "stuck" agents approaches 0, the term goes to ∞ , and unanimity prevails. With unanimity the model can only produce a small finite number of stationary equilibria and there is little gain from complicating the basic multiple equilibria models. The primary modeling gain of our approach is the ability to generate a wide range of variances, through the introduction of a portion of the population that is fixed.

An Aside on Voter Models and Unanimity

The reader familiar with the literature on voter models (e.g., Kindermann and Snell (1980)), will recognize that except for the presence of some agents who are "stuck" at their initial action *i.e.* those $i \in S$, this is exactly a one dimensional voter model. Our perturbation of the basic voter model (adding fixed agents) is however important, since standard one dimensional voter models have stationary distributions that exhibit unanimity, which means that the models predict even more variance than we see in the data. This unanimity property is also present if we consider a lattice in two dimensions, where each agent is indexed by a pair of integers (i, j) , and a neighbor of an agent is any agent $(i, j \pm 1)$ or $(i \pm 1, j)$.

Unanimity, however, disappears in higher dimensions. In dimension three, *i.e.* where each agent is indexed by a triple (i, j, k) , non-degenerate stationary distributions exist. Under any invariant distribution, the number of agents that chooses action one in a cube of population n ,²⁴ (when divided by $n^{5/6}$) converges to a normal random variable (c.f. Bransom and Griffeth (1979)). The constant of normalization here is $n^{5/6}$ as opposed to the usual square root

²³A good approximation for $f(\pi)$, especially for values of $\pi > .025$, is $f(\pi) = 1.88/\pi - .88$.

²⁴The cube, therefore, has size $n^{1/3}$.

of the number of random variables considered (which would be $n^{1/2}$ in this case). In our section on empirical results, we will show that this three-dimensional voter model predicts, like the one and two dimensional models without fixed agents, an even higher variance than that found in the data.²⁵

It is well known that voter models have a very distinct behavior on lattices of dimension one or two *vis-a-vis* lattices of dimension three or greater. Numerical simulations in dimension two and three seem to show that in our modified model the behavior is similar for dimension one, two and three. However, our analytic calculations exposted above were done only for the one-dimensional case.

One difference between models of one or two dimensions and models of three dimensions is the predicted variance of crime rates across cities. For one or two dimensional models (without fixed agents), as n grows large, per capita crime rates still display positive variance (because these models predict unanimity). In dimension three (without fixed agents), as n grows, the variance of per capita crime rates go to zero (as they do with fixed agents in the one-dimensional case). Thus, the number of dimensions can determine the predicted variance across cities, and dimensionality provide an alternative method (to fixed agents) of capturing greater or lesser degrees of interaction and urban heterogeneity.²⁶ While, we see in Table 2 that models (without fixed agents) of dimensions one, two or three predict too much variance to fit the data, it is possible that a model of some higher dimension would generate the observed crime rates. However, the mathematics of such a model would be much more complicated as little is known about voter models in dimensions greater than three.

²⁵The one and two dimensional voter models also predict too high of a variance when there are no fixed agents.

²⁶It is worth noting the strange result that higher dimension models have less variance, not more, because they have more interaction. Low dimension models converge on a unanimous outcome which produces the highest variance across cities. In three dimensions, unanimity is avoided (because of the large number of interactions) and in fact the cross-city variance declines.

In the next section, we discuss a model with explicitly maximizing agents and argue that the stationary version of this first model of local interactions can be interpreted as a particular version of an optimizing model.

Model Two -- Maximizing Behavior and Social Interactions

In the previous model, agents' decisions were not explicitly derived from a maximizing model. Here we present an alternative model with maximizing actors that will generate the same statistical behavior as the first model. Unfortunately, the cost of introducing maximization is that we are forced to assume here each agent's imitation is unidirectional, i.e. when agents are arranged on a line, each agent is influenced only by the agent on their right not by the agent on their left. The modeling difficulties inherent with mutual imitation (which we will comment on later) make this assumption necessary.²⁷

As before, each agent is indexed by an integer i , each agent can choose one of two actions $\{0,1\}$. Action 1 will be interpreted as committing a crime. The utility of an agent i , depends on the action that he takes and also the action taken by agent $i-1$.²⁸

There are three types of agents in this model, indexed 0, 1 or 2. If an agent is of type $\tau \in \{0,1,2\}$, we write his utility as $U_\tau(a_i, a_{i-1})$. If an agent is of type 0 his utility is maximized at $a_\tau = 0$, independent of $a_{\tau-1}$. If an agent is of type 1 his utility is maximized at $a_\tau = 1$, independent of $a_{\tau-1}$. These two types of agents are diehard lawbreakers (type 0 agents) and law abiders (type 1 agents); they are far from the margin and will be uninfluenced by the actions of their neighbor. These two types are analogous to the fixed agents in the previous model.

²⁷ We would prefer to be able to model bi-directional imitation.

²⁸ This interconnection of utilities can be interpreted with any of our positive interaction models. When agent $i-1$'s criminal activity influences the arrest rate, or the return to legal activities, it influences the utility of agent i and the utility agent i gets from criminal activities. For example, agent i receive more utility from committing a crime if agent $i-1$ is a criminal because agent i has learned how to steal more effectively from watching agent $i-1$.

For Type 2 agents, we assume that $U_2(1,1) > U_2(1,0)$ and $U_2(0,0) > U_2(0,1)$, i.e. a Type 2 agent prefers to imitate his predecessor. This type of agent is marginal enough in his decision to become a criminal that he will be dependent on his neighbor's decision. Each agent i is of type 0 with probability p_0 and of type 1 with probability p_1 , and since p_0+p_1 is the total proportion of non-marginal (fixed) agents we will occasionally refer to p_0+p_1 as π here as well. The probability of becoming type 0, 1 or 2 is independent across agents. The most appealing interpretation of type is that each agent is endowed with a different amount of a continuous attribute that determines that net benefit from crime and that the types simply reflect cutoff points in this continuously distributed attribute.²⁹

Each agent observes the action chosen by his predecessor. There is a unique Nash equilibrium given the types of each agent: all strings of type 2 agents uninterrupted by type 0 or type 1 agents will imitate the action of the type 1 or type 0 agent that began the string. The remainder of this study examines the distribution of these Nash equilibria across cities.

If each agent's action $\{a_i\}$ is thought of as a random variable that assumes the values of 0 or 1, then the process $\{a_i, -\infty < i < \infty\}$ is stationary. The presence of agents of type 0 and 1 guarantees enough "mixing" so that the central limit behavior can be established. More precisely, we know that agent i and agent j are independent conditional on the existence of an agent of type 0 or 1 in the segment (i,j) . The probability that there is no such agent in that interval goes to zero exponentially as $i-j$ goes to infinity. This fact gives us sufficient conditions for central limit behavior.

²⁹This continuous attribute could be the returns to a legal, alternative activity or the suffering incurred by being arrested or the ability to commit crime or some weighted combination of all of these attributes. Hence, the cutoff points, and the probabilities p_0 and p_1 , may depend on some city characteristics. In fact, urban characteristics will jointly determine p and π . For example higher employment levels will induce more agents who are always law abiders and fewer agents who are always law breakers, since the opportunity cost of crime has risen. The population percentage p will fall. However, the number of influenceable agents may rise or fall depending on whether the rise in law abiders (who were presumably formally influenceable agents) outweighs the fall in law breakers (who will not become influenceable agents).

Following standard reasoning, consider the set of all agents $|i| \leq n$. The expected number of crimes *per capita* by these agents is denoted as $p \equiv p_1 / (p_0 + p_1)$. Hence, again it must be true that³⁰:

$$\frac{1}{\sqrt{2n+1}} \sum_{|i| \leq n} (a_i - p) \longrightarrow N[0, \sigma^2], \quad (3.1)$$

and further, that,

$$\sigma^2 = \text{var}(a_0) + 2 \sum_{i>0} \text{cov}(a_0, a_i) \quad (3.2)$$

Since a_0 follows a binomial distribution, $\text{var}(a_0) = p(1-p)$.

To compute the covariance terms in (3.2), let A be the event that at least one agent in the segment $[0, i]$ is of the type 0 or 1. We know that conditional on A , a_i is independent of a_0 (since a_i is determined exclusively by the value taken by the agent of type 0 or 1 located between 0 and i). The probability that A *does not* occur is given by $(1 - p_0 - p_1)^i$. If A does not occur, then $a_i = a_0$.

Using this observation, we can rewrite (3.2):

$$\sigma^2 = p(1-p) + 2 \sum_{i>0} p(1-p)(1-p_0-p_1)^i = p(1-p) \left[1 + 2 \frac{1-p_0-p_1}{p_0+p_1} \right] = \frac{p(1-p)(2-\pi)}{\pi} \quad (3.2')$$

The term $2(1-p_0-p_1)/(p_0+p_1)$, represents the covariance across agents and results from the presence of imitation. If no type 2 agents are present ($p_0+p_1=1$), this term is null. As the probability that each agent is of type 2 approaches 1, (i.e. π approaches zero) this term approaches ∞ . Moreover, we know that because the covariances are decaying exponentially the variance of the left hand side of (3.2') converges to σ^2 at the rate at least $1/n$.

If we chose to write $f(\pi) = (2-\pi)/\pi$, the parallel between this solution and the solution of the previous model becomes even clearer (except that now we have a closed form solution for $f(\pi)$). With this model, as with the previous model, the variance of crime rates (times the square root of population) across

³⁰ Again see White (1984), Theorem 5.19, page 124.

cities equals $p(1-p)$ times a function of the proportion of the population that does not react to outside influences. Figure 1 graphs this explicit functional form for $f(\pi)$ with numerical values found for $f(\pi)$ in the previous model. As the figure shows, the two different models yield almost identical estimates for $f(\pi)$ except at low (below .05) values of π .³¹

In this model it is easy to transform the estimated $f(\pi)$ into an estimate of the average size of an interdependent social group. As in the previous section the average social group size is $1/\pi$. Using the new formula for $f(\pi)$, we find that our estimate of the size of the social group is $(f(\pi)+1)/2$.

An Aside on Symmetric Imitation

If we wish to introduce more symmetric imitation *i.e.* the utility function of agent i depends on the actions of agents $i-1$ and $i+1$, one natural assumption is to assume that agents utilities depend on the sum of their neighbors' actions. So we can write $s_i = a_{i-1} + a_{i+1}$, and assume that the utility of agent i is given by $u_{\tau_i}(a_i, s_i)$ where τ_i is his type. In a bilateral model, we again assume the existence of two fixed types of agents (again denoted type zero and one), the utility of type 0 is maximized by action zero, and that of type 1 by action 1, independently of s_i . If we imagine as before that there is only one other type, than it is natural to assume that for that type $u_{\tau}(0,0) > u_{\tau}(1,0)$, $u_{\tau}(1,2) > u_{\tau}(0,2)$, and, for symmetry, $u_{\tau}(0,1) = u_{\tau}(1,1)$. In this case we are back exactly to the first model!

If, however, we try to dispense with ties, it is necessary, again for reason of symmetry, to use four types. Types 0 and 1 are as before, and both type $\tau = 3$ and $\tau = 4$ satisfy $u_{\tau}(1,2) > u_{\tau}(0,2)$ and $u_{\tau}(0,0) > u_{\tau}(1,0)$. Furthermore, $u_3(0,1) > u_3(1,1)$ and $u_4(1,1) > u_4(0,1)$, *i.e.* when neighbors disagree, type 3 is more inclined towards action 0 and type 4 towards action 1. In this case, however, it is easy to see that there are a large number of Nash equilibria.

³¹In fact, our estimates of π will be in this low value region, so in fact, the choice of model will determine significantly our estimate of π . As a result, we believe it is reasonable to discuss primarily the estimates of the $f(\pi)$ variables and not emphasize too strongly the actual estimates of π .

To see what can happen, take an interval consisting of types $\{0, 4, 3, 1\}$. It is obvious that outcomes $\{0, 0, 0, 1\}$ and $\{0, 1, 1, 1\}$ are both Nash equilibria (in fact there are many others). Because of the mutual positive interactions this represents an exceedingly complicated coordination game where no simple refinements apply and there is no way to satisfactorily characterize predicted behavior. As a result we will simply avoid this more difficult model.

4. Empirical Framework

This framework is meant to discuss the issues involved in empirically estimating $f(\pi)$. In particular, we are interested in allowing for urban heterogeneity in the costs and benefits of crime. The probability that any single citizen in city j chooses to become a criminal is denoted $p_j^* = p^*(z_j, \epsilon_j)$, where the variables in the vector z_j are observable urban characteristics (population of the city, levels of urban schooling, local labor markets, etc.), and the variables in the vector ϵ_j are unobservable urban characteristics.³² Urban characteristics predict an individual's propensity to commit a crime because they reflect the individual's own characteristics: an individual from a high schooling city is, by definition, more likely to have high schooling. More schooling may lower the propensity towards crime by raising the cost of crime. Urban characteristics also determine the environment in which the agent lives which may also affect the propensity for crime. It may be more difficult or costly to commit crimes because the populace is well educated. Aggregate data do not present us with any way to disentangle these two effects so we will ignore the fact that local area characteristics affect crime rates for two very different reasons.³³

Our empirical work takes agents' decisions about migration as predetermined. If criminals choose to live with other criminals, then our results may overstate the effect of local influence on criminal behavior (and

³²It is certainly true that individual characteristics change the likelihood of being a criminal far more than local area characteristics do, and this model is not inconsistent with that statement. The $p^*(\cdot, \cdot)$ function gives us the probability that the agent is a criminal based only on our information that the individual lives in town j . The individual information determines whether this probability develops into a criminal outcome.

³³Case and Katz (1992) do a better job at disentangling the two effects with individual level data, which is necessary to distinguish between these two different effects of urban characteristics.

understate the role of endogenous migration). We will return to this question later, but the high percentage of criminals who are youths makes it obvious that the majority of criminals do not have the opportunity to make their own migration decisions (according to the 1991 Uniform Crime Reports more than 50% of those arrested for property crimes are 21 or under). Furthermore, locational sorting is itself evidence of positive interactions; if criminal choice to live together, then there must be positive interactions.

We will let p_j denote the actual amount of crime in each city, where $p_j = p(z_j, \varepsilon_j, \omega_j)$. The term ω_j is meant to reflect the random shocks that may make a city's crime rate differ from the crime rate that would be predicted knowing all of the city's basic attributes. We assume that ω_j is independent of (z_j, ε_j) .

This estimation framework is meant to decompose the gap between actual crime rates and the crime rates that are predicted using observable characteristics. The decomposition divides this gap into a portion related to imperfect observation of the city's characteristics (the vector ε_j) and a portion related to agents' deviating from their expected level of crime (the ω_j vector). This decomposition will allow us to estimate the amount of social interactions, since the quantity of social interactions (more precisely $f(\pi)$) has been found in the previous section to determine how far a city's crime rate will differ from the crime rate that would be predicted if all of the city's characteristics were known.

Our basic strategy for the decomposition, which will enable us to estimate the size of the interactions, is to assume a structure for the "error term" that comes from missing city-level characteristics. Specifically, we assume that p_j^* , the prediction given complete information on the city's characteristics) has a standard logit functional form:

$$p_j^* = \frac{e^{v_j}}{1 + e^{v_j}} \quad (4.1)$$

where v_j is mean zero, finite variance, and $v_j = \zeta_j + \varepsilon_j$, where both ζ_j and ε_j are also mean zero, finite variance terms. ζ_j is a function of the observables, z_j ,

and ε_j is independent of these observables. If we denote as \hat{p}_j the prediction of the city's crime rate based on z_j then:³⁴

$$p_j^* = \frac{\hat{p}_j e^{\varepsilon_j}}{1 - \hat{p}_j + \hat{p}_j e^{\varepsilon_j}} \quad (4.1')$$

We further assume that ε_j is a normally distributed error term with $E(\varepsilon_j) = 0$, and $\text{Var}(\varepsilon_j) = \lambda^2$. As we showed earlier, with either model of social interactions:

$$\text{var}(p_j - p_j^*) \sqrt{N_j} \rightarrow f(\pi) p_j^* (1 - p_j^*), \quad (4.2)$$

and this convergence is at rate $1/N_j$.

From now on we take the vectors z_j as fixed and all random variables indexed by j will be considered as functions of $(\varepsilon_j, \omega_j)$. We will denote:

$$y_j \equiv (p_j - \hat{p}_j) \sqrt{N_j} = [(p_j - p_j^*) + (p_j^* - \hat{p}_j)] \sqrt{N_j} \quad (4.3)$$

Taking expectations conditioning on ε_j yields:

$$E(y_j | \varepsilon_j) = E[(p_j - p_j^*) \sqrt{N_j} | \varepsilon_j] + E[(p_j^* - \hat{p}_j) \sqrt{N_j} | \varepsilon_j] \quad (4.4)$$

Using the facts that $E(p_j | \varepsilon_j) = p_j^*$ and that $E(p_j^* | \varepsilon_j) = \frac{\hat{p}_j}{1 + \hat{p}_j (e^{\varepsilon_j} - 1)}$:

$$E(y_j | \varepsilon_j) = \frac{\hat{p}_j (1 - \hat{p}_j) (e^{\varepsilon_j} - 1)}{1 + \hat{p}_j (e^{\varepsilon_j} - 1)} \sqrt{N_j}. \quad (4.4')$$

Again taking expectations conditioning on ε_j yields:

$$\text{Var}(y_j | \varepsilon_j) = \text{Var}[(p_j - p_j^*) \sqrt{N_j} | \varepsilon_j] + \text{Var}[(p_j^* - \hat{p}_j) \sqrt{N_j} | \varepsilon_j] = \text{Var}[(p_j - p_j^*) \sqrt{N_j} | \varepsilon_j] \quad (4.5)$$

³⁴Of course $p_j^* = \frac{e^{\varepsilon_j}}{1 + e^{\varepsilon_j}}$

since $(p_j^* - \hat{p}_j)$ is a constant (and thus independent of ω_j) when we condition on ε_j . Using (4.5), (4.2) and the results of either model, (4.5) can be rewritten:

$$\text{Var}(y_j|\varepsilon_j) = \text{Var}[(p_j - p_j^*)\sqrt{N_j}|\varepsilon_j] = f(\pi)p_j^*(1 - p_j^*) + O(N_j^{-1}) \quad (4.5')$$

where $O(N_j^{-1})$ refers to terms on the order of $1/N_j$ which we shall ignore.

Using a standard result (see, e.g. Feller (1968)), it follows that:

$$\text{Var}(y_j) = \text{Var}[E(y_j|\varepsilon_j)] + E[\text{Var}(y_j|\varepsilon_j)] \quad (4.6)$$

Since \hat{p}_j and $\sqrt{N_j}$ are deterministic functions of the z_j vector, we know that:

$$\text{Var}[E(y_j|\varepsilon_j)] = \text{Var}\left[\frac{\hat{p}_j(1 - \hat{p}_j)(e^{\varepsilon_j} - 1)}{1 + \hat{p}_j(e^{\varepsilon_j} - 1)}\sqrt{N_j}\right] = \hat{p}_j^2(1 - \hat{p}_j)^2 N_j \text{Var}\left[\frac{(e^{\varepsilon_j} - 1)}{1 + \hat{p}_j(e^{\varepsilon_j} - 1)}\right] \quad (4.7)$$

Using (4.5'):

$$E[\text{Var}(y_j|\varepsilon_j)] = f(\pi)E[p_j^* - p_j^{*2}] = f(\pi)[\hat{p}_j - \hat{p}_j^2 - E(p_j^* - \hat{p}_j)^2] \quad (4.8)$$

Combining (4.7) and (4.8) and using the fact that $E(p_j^* - \hat{p}_j)^2 = \frac{\text{Var}(y_j|\varepsilon_j)}{N_j}$,

$$\text{Var}(y_j) = (N_j - f(\pi))\hat{p}_j^2(1 - \hat{p}_j)^2 \text{Var}\left[\frac{(e^{\varepsilon_j} - 1)}{1 + \hat{p}_j(e^{\varepsilon_j} - 1)}\right] + f(\pi)[\hat{p}_j - \hat{p}_j^2] \quad (4.6')$$

If we assume that ε_j is normally distributed then there exists a function that maps the standard deviation of ε_j (denoted λ) and \hat{p}_j into the variance term in (4.6'). Specifically, we will denote:

$$\Psi(\lambda, \hat{p}_j) = \text{Var}\left[\frac{(e^{\varepsilon_j} - 1)}{1 + \hat{p}_j(e^{\varepsilon_j} - 1)}\right] \quad (4.9),$$

For empirical implementation we can define:

$$\mu_j \equiv (p_j - \hat{p}_j)^2 N_j - E[(p_j - \hat{p}_j)^2 N_j] = (p_j - \hat{p}_j)^2 N_j - \text{Var}(y_j) \quad (4.9)$$

where μ_j will be mean zero and will reflect the extent to which we do not perfectly observe the variance of crime levels. Using (4.6'),(4.9) and (4.10):

$$(p_j - \hat{p}_j)^2 N_j = \hat{p}_j^2 (1 - \hat{p}_j)^2 (N_j - f(\pi)) \Psi(\lambda, \hat{p}_j) + f(\pi) \hat{p}_j (1 - \hat{p}_j) + \mu_j \quad (4.10)$$

This equation can be estimated when we have an approximation for $\Psi(\lambda, \hat{p}_j)$.

In particular, when we know that $\hat{p}_j = p$, i.e. the expectation of crime is constant across cities and is equal to the aggregate average:

$$(p_j - p)^2 N_j = p^2 (1 - p)^2 (N_j - f(\pi)) \Psi(\lambda, p) + f(\pi) p (1 - p) + \mu_j \quad (4.10')$$

Estimating $f(\pi)$ to minimize $\sum \mu_j^2$, and taking p and λ as known:

$$\hat{f}(\pi) = \frac{\sum_j [(p_j - p)^2 N_j - p^2 (1 - p)^2 N_j \Psi(\lambda, p)]}{J[p(1 - p) - p^2 (1 - p)^2 \Psi(\lambda, p)]} \quad (4.11)$$

where $\hat{f}(\pi)$ denotes our estimate of $f(\pi)$. Equation (4.11) makes it obvious that different estimates of λ will result in different estimates of the amount of social interaction. In cases, where \hat{p}_j 's are known, as are λ 's, (4.11) becomes:

$$\hat{f}(\pi) = \frac{\sum_j [(p_j - \hat{p}_j)^2 N_j - \hat{p}_j^2 (1 - \hat{p}_j)^2 N_j \Psi(\lambda, \hat{p}_j)]}{\sum_j [\hat{p}_j - \hat{p}_j^2 - \hat{p}_j^2 (1 - \hat{p}_j)^2 \Psi(\lambda, p)]} \quad (4.11')$$

We can estimate λ fitting equation (4.10) and we can find estimates of both λ and $f(\pi)$ simultaneously. In that case, our estimates of λ hinge on the correlation of N_j and $(p_j - \hat{p}_j)^2 N_j$ (which comes from the fact that expectation of the variance of y is independent of N but that the variance of the expectation of y rises with N).³⁵

³⁵The critical assumption for this estimation procedure to work is that λ be essentially independent of N .

So far in our discussion, we have assumed that $\Psi(\lambda,p)$ is known (even if λ is not). In fact, no closed form solution for $\Psi(\lambda,p)$ to our knowledge exists. To solve this problem we used simulation data and found a non-linear approximation for $\Psi(\lambda,p)$. Our simulations involved fixing 4,000 (λ,p) pairs and then generating 10,000 values of ε_j for each (λ,p) pair. Given these generated ε_j 's, we then estimated $\Psi(\lambda,p)$ for each (λ,p) pair.³⁶ We fit our sequence of $\Psi(\lambda,p)$'s by ordinary least squares to a polynomial containing (λ,p) terms-- this regression had an R^2 of over 99%. We use this polynomial approximation of $\Psi(\lambda,p)$ throughout the empirical work.

5. Results

The data used in our estimation come from two sources: the FBI and the New York City Police Department. The cross-city crime data are published by the FBI under the *Uniform Crime Reporting* program. The data are compiled from monthly reports submitted to the FBI by over 16,000 city, county, and state law enforcement agencies. Our New York City data comes from 1993 reports of crimes by precinct. Both data sets detail crimes reported (and verified), rather than arrest or survey data.

Obviously, these data may undercount the actual number of crimes committed since many crimes are not reported. Our suspicion is that the amount of under reporting will be higher in high crime areas, so that cross-city variance is underestimated, but it is possible that under reporting is greater in those areas without much crime. In that latter case, we would observe a spuriously high cross-city (or cross-precinct) variance in crime rates.³⁷

The crimes' definitions are included in an Appendix which describes the variables used in the study. We use not only crime rate data but also data from City and County Data Books on urban characteristics. Our unit of

³⁶By experimenting with different sample sizes and by repeating the simulations we are sure that 10,000 observations is enough so that our estimate of $\Psi(\lambda,p)$ (knowing (λ,p)) is quite accurate.

³⁷Levitt (1994) discusses the problems of using reported crimes more thoroughly.

analysis throughout is the city, and the data books conveniently provide us with city-level data (from the census) for both 1970 and 1985 (in some cases we were forced to match 1980 data with 1985 and 1986 FBI data). For precinct level data, we matched the police data with 1990 census data presented in the New York Department of City Planning *1990 Census Data by Police Precinct*. This volume's mapping between precincts and census tracts may be less precise than the exact mapping (between cities and cities) allowed by the census and FBI data.

Tables 1A-1E -- Means, Standard Deviations and the Sample's Extremes

Tables 1A-1E presents a quick look at our data. Table 1A presents the five cities with the most and least serious crimes per capita in 1985. Two of the highest crime cities in the United States owe their position (in part at least) to the Census definition of population which does not include the large tourist populations of Atlantic City and West Palm Beach.³⁸

While the high crime levels of Atlantic City and West Palm Beach are related to their high tourist populations (because of undercounting of population or because tourists attract criminals), Portland, Oregon is an extremely unusual crime center. Portland has modest unemployment and high levels of education. Still, Portland in 1980 had extremely high levels of serious crimes (and has had high levels of crime from 1975 onward). Highland Park's crime rate is more explicable. That city has high levels of unemployment, and extremely high levels of female headed households. Miami is slightly less explicable than Highland Park, but it too has low levels of human capital in its population.

The five least crime ridden communities all share extremely low population levels. This finding is related to the well-known positive correlation between

³⁸ If our goal in using crimes per capita is to reflect the number of crimes relative to the population of potential victims, then as the measured populations of West Palm Beach and Atlantic City fail to include their large tourist populations, then the measured crimes per capita in those cities overstates the ratio of crimes to potential victims. If, instead, we are interested in measuring the ratio of crimes to potential criminals (as we are in this study) as long as the criminals come from among the long term residents (not the tourists), then the census definition of population is appropriate.

city-size and crime level.³⁹ Most of these cities all have fairly high levels of human capital and low levels of unemployment. The one exception to this statement is Garfield, New Jersey, which manages to have the second lowest crime rate with only 44.9% of its citizens being high school graduates.

Tables 1B-1E present means and standard deviations of our data sets. Table 1B presents the means and standard deviations for 1985 cross-city data. The first line of the data series shows that there are on average .067 serious crime per capita per year in the U.S. More strikingly, the city with highest serious crime rate has 62.9 times more serious crimes per capita than the city with the lowest crime rate. For all of the other crime groupings, the minimum per capita level is zero. The maximum crime rate for an individual crime is .2968 crimes per capita for larcenies.

Throughout this paper, when we examine individual crimes we will treat these crimes as independent social phenomena, i.e. we will not examine the whether higher murder rates relate higher levels of arson. Naturally, our assumption of independence is incorrect in many cases and there is surely a great deal of information to be gained by examining interactions across crimes. When we examine serious crimes generally, we make the opposite assumption and assume that criminal interactions work as strongly across crimes as they do within crimes. Again, this extreme is also unrealistic and we hope to create a more flexible mechanism for dealing with cross-crime interactions in later work.

While the cross-city differences in crime rates are remarkable, there are also dramatic cross-city differences in schooling, government expenditures and almost every urban attribute. One response to the observed heterogeneity of crime rates across cities is to belittle all of the cross city variance in attributes and see them as "normal." Our reaction is to argue that this high degree of variance suggests interactions in other variables besides crime and that the

³⁹ This result may, in principle, also be related to cities with lower population levels having more idiosyncratic features that we do not include in our basic regressions. That argument suggests that it is more likely to see smaller cities at the tail of the distribution. Our later works suggests that empirically there is a positive link between crime rate variance and city population, so that this explanation for five small cities being at the tail of the distribution is unacceptable.

previous models can be useful for examining social interactions in other spheres as well (such as education⁴⁰). We will, in fact, attempt to use our methodology briefly for non-criminal variables.

Table 1C gives cross city data for 1970. The mean level of reported crimes has more than doubled between 1970 and 1985. The standard deviation of crime rates across cities has risen by approximately 50%. The highest crime rate city in 1970 has approximately 20 times more crimes per capita than the lowest crime rate city in 1970.

These numbers make it obvious that we need a model based index of interactions. A naive empiricist might try to use some basic measure of variance to measure the degree of interactions. But if we use standard deviation divided by mean, to measure the degree of interaction we would think that interactions had fallen. At the same time if we used the comparison of the ratio of highest crime rate city to lowest crime rate city it seems as if interactions have risen. It is impossible to choose the appropriate measure of social interactions without a model that gives us a estimable parameter that captures the degree of interaction.

Table 1D provides the cross-precinct data for New York City. The mean level of crimes in New York is approximately the same as the mean level of crimes for the nation. However, the definition of serious crimes differs between our New York data and the cross-city data.⁴¹ For assaults, murders, rapes, robberies and auto thefts, New York is well above the national average. For larcenies and burglaries, New York is below the national average (primarily because of the definitions of these crimes).

Cross precinct variation in New York is quite high. The safest precinct has one tenth the serious crimes per capita of the most crime ridden precinct. The ratio for larcenies between the most crime ridden precinct and the least crime ridden precinct is over 300 to one. Of course, there is also a great deal of variation across precincts in many underlying characteristics as well.

⁴⁰There do in fact exist many models of local interactions outside of crime. Benabou (1994) is an excellent example of interactions in the production of human capital.

⁴¹Fewer crimes are classified as serious in the New York data.

Table 1E gives means and standard deviations for 1986 cross city data. This data is remarkable in that there are substantial differences in national averages between 1985 and 1986 -- auto thefts rose by 15% in a single year. The differences at the city level are even more remarkable. These intertemporal differences suggest further that there may be correlations across individual decisions to engage in crime.⁴²

Tables 2A and 2B-- Basic Results

Table 2A presents our basic results on the presence of social interactions. The table is organized by crime beginning with the aggregate measure -- serious crimes, which is the sum of the other crime variables. Within each crime category we have evidence from the three data sources. The columns marked 1985 refer to the FBI cross-city data in 1985. The columns marked 1970 refer to the same FBI cross-city data but for 1970. The columns marked NYC refer to 1993 data for New York City precincts. While the number of locations is much smaller for NYC precincts (70 instead of 688), the average population of the precincts is roughly equivalent to the average population of cities in the cross-city sample. In Table 2A the sample size for 1985 cross-city data and NYC data do not change across columns. The sample size for 1970 cross-city data drops three times to a smaller sample of 541 cities because of data availability (for murder, rape and larceny data).

The first column gives the average U.S. crime rate times one minus the average crime rate for each row. This number is the variance of cross location crime rates that would be expected if criminal decisions were independent and if the expected proportion of criminals was constant across cities. This number provides us with a benchmark predicted variance.⁴³

Since we are estimating the variance across cities from a finite sample of cities, we are not surprised when the observed variance differs from the

⁴²A high variance in aggregate crime levels would require correlations across cities as well as across individuals.

⁴³ More accurately, this number, $p(1-p)$, is the predicted variance of cross city crime rates minus U.S. average times the square root of city population.

predicted variance. To gain confidence intervals around the predicted variance, we will use the fact that a sum of squared, standardized normal random variables follows a chi-squared distribution. Using this fact, and a standard chi-squared distribution table, we know that the observed variance of 70 independent units will lie between .61 and 1.49 times the true variance ($p(1-p)$) 99% of the time. Higher group sizes yield smaller confidence intervals. So an observed variance more than 1.5 times the predicted variance allows us to reject the null hypothesis. Since the observed variance is often over 1000 times the predicted variance and never less than twice the predicted variance, we will refrain from further discussion of statistical significance throughout the bulk of the remaining results.

The second column of Table 2A gives us the actual variance of crime rates (times the square root of city population) across locations. The difference in magnitude between the first and second columns is staggering for many of the rows. The third column gives our estimate of $f(\pi)$, which when $\lambda=0$ is the ratio of the second column to the first column. This ratio moves from a low of 10 (for murders) to a high of 1313.8 for serious crimes generally. There are several ratios of over 1000.

The high values of $f(\pi)$ for serious crimes, relative to many other crimes, is explained by two distinct forces. First, the bulk of serious crimes are larcenies and $f(\pi)$ values for larceny are extremely high. Second, the observed $f(\pi)$ for serious crimes is a function both of (1) the interactions within crimes and (2) the interaction across individual crimes (i.e. the fact that more murders decrease the costs of robbery). Since, the $f(\pi)$ for serious crimes includes both of these effects, while the $f(\pi)$'s for individual crimes includes only within crime interactions, we would expect the $f(\pi)$ for serious crimes generally to be higher than a weighted average of the $f(\pi)$'s estimated for the individual crimes. While, we are not seriously examining the interaction across crimes here, this high level of $f(\pi)$ suggests that there is a significant quantity of inter-crime spillovers.

The preceding analysis essentially assumes that there are no underlying differences across communities in the propensity to engage in crimes. The next two columns both allow heterogeneity in underlying community

characteristics. As discussed in the estimation framework section, as a preliminary step before estimating λ , we parametrize this heterogeneity by assuming two extreme values of λ which denotes the size of attribute differences across cities. We use two potential values for λ : .004 and .04. The lower estimate of λ corresponds to the amount of observed heterogeneity in observables. In the model, $p^* = \frac{\hat{p}e^{\lambda}}{1 - \hat{p} + \hat{p}e^{\lambda}}$. If we assume that p^* is the crime rate predicted by a city's observable attributes⁴⁴ and \hat{p} is a constant across cities, then we can estimate λ . Our estimate of λ found using this method, which we feel is something of a reasonable lower bound, is .004. Our upper bound, .04, was chosen (1) because it represents 10 times more urban heterogeneity than we actually observe in the data and (2) it is significantly larger than any of the estimated values of λ that we find later in the paper. We then reestimate $f(\pi)$, allowing the variance of underlying propensities to commit crimes to be that predicted with these different λ values. The results from these reestimations are in columns 5 and 6 of Table 2A.

When we allow the variance of underlying propensities to commit crimes to vary across cities to be non-zero, the results are basically unchanged for most of the rows. Introducing this correction affects the results most for those crimes that occur most often, because these corrections essentially deflate $f(\pi)$ by a number proportional to the average level of crime. Thus, the serious crime figure is most affected because the raw level of serious crimes is much higher than in the other variables. The other crime levels are much less affected. Nevertheless, it is obvious that $f(\pi)$'s are still wildly higher than one. Furthermore, the basic levels of $f(\pi)$ do not change too much. The estimated $f(\pi)$ for serious crimes moves from 1313.8 to 941.4 as we move λ from 0 to 0.04. The estimated $f(\pi)$ for murder moves from 10 to 9.5 as we move λ from 0 to 0.04.

The rankings of crimes by social interactions seems to be relatively robust to the value of λ . For 1985, $\lambda=0$, the ranking of crimes by social interaction is: serious crimes, larceny, auto theft, burglary, robbery, assault, arson, rape and murder. For 1985, $\lambda=0.04$, the ranking of crimes by social interaction is:

⁴⁴We used a standard logit estimation procedure.

serious crimes, auto theft, larceny, robbery, burglary, assault, arson, rape and murder. The correction for city-level attribute differences changes the order of auto theft and larceny and robbery and burglary, but basically these lists are very similar.

For 1970 data when $\lambda=0.04$, the ranking of crimes by social interaction is: serious crimes, auto theft, robbery, burglary, larceny, assault, rape and murder (we have no arson numbers). Using 1970 numbers, only changed the order of larceny. Figure 5 shows the correlation between $f(\pi)$ estimates for 1985 and 1970.

For New York city, when $\lambda=0.04$, the ranking of crimes by social interaction is: larceny, serious crimes, robbery, auto theft, assault, burglary, murder and rape. Figure 6 shows the correlation between $f(\pi)$ estimates from 1985 cross-city data and from within New York data. The New York data has some difference in its ranking of crimes by social interactions, but the magnitudes of the estimated $f(\pi)$'s all remain the same (except for assault),⁴⁵ and given the severe differences between (1) New York and the rest of the country and (2) the possibility for cross-district interactions (which would make comparison between cross-district and cross-city data difficult), we are more surprised by the similarities between the New York data and the cross-city data than we are by any differences. In fact, we take the remarkable similarities between estimates based on cross-city and estimates based on cross-precinct data as evidence that many of the interactions that take place within a city take place within the district, so that the appropriate interaction mechanisms must be more local, and less aggregate, in character.

The next column gives the λ value necessary for us to believe that no covariance across agents exists, which we denote as λ^* . These λ^* value can be seen as an alternative measure of how strong social interactions are in a particular crime. Higher λ^* values mean that it is less plausible that the observed level of intercity variance can be rationalized by differences in the

⁴⁵ The New York data shows substantially lower levels of interactions in assault than our cross-city data. One explanation for this discrepancy is in the reporting or classification of assaults. A second explanation is that there are cross-district interactions that show up in the cross-city data but do not appear in the cross-district data.

underlying propensity to commit crimes. Lower λ^* values make it more likely that cross-city differences may just be differences in economic conditions. These λ^* values seem to us quite high.

The final column in Table 2A gives the variance of city crime rates weighted by population to the 5/6 not by population to the 1/2. The prediction of 3-dimensional voter models (as discussed earlier) is that this number will be close to the predicted cross-city variance, $p(1-p)$. In fact, this variance is usually far too low. The three dimensional voter models predict more cross-city crime rate variance than we observe.⁴⁶

Table 2B repeats the columns of Table 2A but for non-crime variables: percentage female in a location and high school graduation rates in 1980 and 1970. We believed that percentage female was a good test of our methodology. Individuals do not choose their sex and it is obvious that if the model showed a high level of interdependence the model was not accurately measuring the level of social interactions. In fact, using the uncorrected $f(\pi)$ measure, we do find a large number of interactions in percent female, but once we allow even the slightest difference in communities ability to attract women this interdependence disappears. The λ^* for women was .001, as opposed to .14 for serious crimes. As soon as we even slightly correct for heterogeneity across cities, our $f(\pi)$ measure yields the expected result: the likelihood of being female in a population does not seem to rise as others with the proportion female in one's neighborhood.

The variable, percentage of the population over age 25 that has graduated from high school ("high school graduates"), seems to display more social interactions than the percentage female variable. The level of interactions among high school graduates in 1980 are roughly comparable to the interactions among larceny perpetrators: $f(\pi)$ is 1149.2, slightly higher than the figure for robbery. However, the λ^* for high school graduates in 1980 is lower than the estimated λ^* 's for all of our crime variables. Our estimate of $f(\pi)$ for education substantially declined between 1970 and 1980 as the cross-city

⁴⁶ It is possible that higher dimension models will match the observable variance, but as mentioned earlier little is known about models of dimension greater than three.

variance in high school graduates declined. Poorer cities (particularly the South) acquired much more education over those years. We believe that these numbers are merely suggestive of a method of producing alternative evidence on education-based spillovers in location or production of education.⁴⁷

Table 3 -- Repeat Offenders

One of the simplifications in Table 2A is that we assume that each crime is perpetrated by a different individual and reflects an agent's decision to enter a life of crime (at least for that year). Agents frequently commit more than one crime within a given year. These repeat crimes provide a degree of correlation across crimes, because if an agent commits one crime, we expect that he will commit more crimes. This effect would reveal itself as a social interaction in Table 2A, that does not actually require any cross-individual interactions. Of course, if we actually had data on criminals rather than crimes this problem would be trivially handled, but since crime data is the only data available it is necessary that we deflate our measures to allow for the possibility that when someone is seen committing one crime it is more likely that the individual will commit more crimes.

More complicated and realistic models of multi-crime criminals are beyond the scope of this paper, but we can make a simple correction assuming that each criminal performs a fixed levels of crimes. Essentially this correction means that we assume that p crimes in a city means that the city has p/R criminals where R refers to the number of repeat offenses per criminal. This correction uniformly lowers the estimated $f(\pi)$'s in our sample. The intuition of this is simple; the adjustment roughly lowers $p(1-p)$ by R (since the p^2 term is negligible here), but it lowers the observed variance by R^2 . As a result, the observed variance and the predicted variance become much closer, and $f(\pi)$ decreases in magnitude.

⁴⁷In education, social interactions could be occurring because as individuals become more educated it becomes cheaper to acquire education (as in Benabou (1994)) or because educated individuals like to live together. Disentangling the interactions that change the migration of the educated and the interactions that determine educational must be left for later work.

Our estimates for the number of crimes per criminal come from various sources. For the serious crimes, we took the value of 6.4 crimes per criminal from a self-reported measure in the Rand Prison Inmate Survey. This number, undoubtedly, is biased upwards since repeat criminals (we believe) are more likely to be incarcerated.⁴⁸ For serious crimes we also estimated $f(\pi)$'s assuming 3 crimes per criminal and 10 crimes per criminal, to find a range of possible estimates for $f(\pi)$. For individual crimes, we used Blumstein and Cohen's study of arrest records. This data may be underestimated since only crimes where arrests were made are counted in the study; there may also be overestimates since only arrested criminals are included in the study (and arrested criminals are more likely to be repeat offenders).

The overall effect of these controls is to lower the estimated $f(\pi)$'s. Still the smallest $f(\pi)$ (with the most conservative λ) is still 80.9, which indicates a great deal of social interaction. These estimates are difficult to use given our lack of knowledge about the structure of repeat offenders, and we take these estimates as extreme lower bounds, but we do know that controlling for repeat offenses fails to eliminate the huge variances across cities and that most of our qualitative results are unchanged by controlling for crimes per criminal. The ranking of crimes by the observed level of social interaction (at any level of λ) is: serious crimes (at 3 crimes per criminal), auto theft, assault, robbery, burglary, and larceny. The primary changes in ranking are that interactions in larcenies fall because of the high level of repeat offenders and assaults show higher levels of social interaction.

Table 4 -- City Specific Characteristics

The previous tables have estimated $f(\pi)$'s using a postulated variance of underlying city characteristics. Table 4 actually controls for a wide range of urban characteristics and then looks at the variation left once the crime rates have been orthogonalized with respect to the urban characteristics described in the appendix.⁴⁹ We included a variety of variables and we believe that the endogeneity of many of these variables should mean that we have overcorrected for city characteristics and that the variance in the fourth

⁴⁸Although it is possible that the really repeat offenders avoid incarceration completely.

column of Table 4 may be an underestimate of the true variance.⁵⁰ Again, we are trying to get an upper bound on the level of social interactions.

We again use our three basic data sets. We have included here also a row for 1986 cross-city data. For this data, we have included the 1985 crime rate of the same city as an explanatory variable. Our view is that including this lagged crime rate, which should eliminate the effect of any omitted urban characteristics, is an extreme check on the presence of social interactions. Since our model is about the choice to become a criminal, and since those choices probably do not change from year to year, controlling for last year's information about who chooses to become a criminal should leave us observing underestimating the true cross-city variance to be explained by local interactions.

The fifth column of Table 4 gives the new estimates of $f(\pi)$ when we orthogonalize crime rates with respect to these urban characteristics. Unsurprisingly, the $f(\pi)$ estimates are uniformly below those of Table 2 (when $\lambda=0$).⁵¹ The numbers for 1970 and 1985 data still reveal the same patterns as Table 2. Figure 8 plots our estimated $f(\pi)$'s in this table against the $f(\pi)$ estimates from Table 2 when $\lambda=0$. The crimes of murder (and to a lesser degree rape) display almost no social interactions.

The 1986 data still reveals substantial social interactions for larceny and serious crimes. Assault, auto theft, and burglary display $f(\pi)$ estimates of over 30 for the 1986 data. Arson, murder, and rape show very little social interactions at all once we control for the lagged crime rates. Our view is that this data helps us conclude that for most of the common crimes, cross-city differences in the propensity towards crime cannot be creating the large cross-city variance in observed crime rates. However, arson, murder, and rape may

⁴⁹We used a standard logit estimation for these predicted urban crime rates.

⁵⁰Whenever crime influences a variable that we included as an explanatory variable, that variable will display explanatory power and will lesson the variance of the residual crime rate. However, the endogenous variable may not actually reflect an underlying difference in urban propensities to crime, rather it reflects the outcome of criminal choices. Including the endogenous variable causes the residual variance to be too low and the predicted variance to be too high.

⁵¹In fact, one can prove that the $f(\pi)$ estimates must fall us urban characteristics are included.

in fact have little substantial social interactions once city level characteristics have been adequately included in the regressions.⁵²

The last two columns of Table 4 go even further. Not only do we use preliminary regressions, and work only with residuals from those regressions, but we also estimate a λ to account for the possibility of omitted city characteristics (we estimate equation (4.10)). Our identification of λ hinges on our model's prediction that (1) the variance of the gap between p^* and \hat{p} is independent of population (so when this variance is weighted by the square root of population, the variance rises with population) while (2) the gap between p and p^* when weighted by the square root of population should be independent of population.⁵³

The estimates of λ range from zero to .027. These estimates support the premise that the variance is not ultimately accounted for by differences in propensities across cities. The estimated $f(\pi)$'s are in the final column and many of the qualitative results of earlier sections are essentially unchanged. We also have standard errors bounding the $f(\pi)$ estimates so we can again note that most of these estimates are different from one. Again serious crimes, burglary, larceny and auto theft are all crime rates displaying a large amount of social interactions. Rape, murder and arson display much lower levels of social interaction.

The ranking of crimes using this estimate of $f(\pi)$ for 1985 data is now: serious crimes, auto theft, larceny, burglary, assault, robbery, arson, rape and murder. The ranking of crimes using our $f(\pi)$ estimate for 1986 (holding 1985 crime rates fixed) is: serious crimes, larceny, auto theft, burglary, assault, robbery,

⁵²As discussed earlier, though, these differences are still statistically distinct from the case where agents' decisions are independent.

⁵³A possible problem with this estimation occurs if λ is itself a function of N . Following a suggestion of Kevin Murphy, we checked for this possibility by using the λ 's associated with the predicted values (i.e. we assumed $p^* = \hat{p}$ and estimated λ by looking at the gap between \hat{p} 's and the mean level of crime). We estimated separate λ values within 10 population groupings of cities and found that these values do not decline with population size within the first eight groups. The two highest groups had somewhat lower λ values. We then re-estimated our $f(\pi)$'s for the smallest 80% of our cities (where we knew that λ value would be constant across populations if observables are distributed like unobservables) and found substantial $f(\pi)$ estimates for that subsample of the data.

arson, rape and murder. Figure 7 shows the strong overall correlation between these estimates of $f(\pi)$ and the earlier estimates

Interpreting $f(\pi)$

In either model, $f(\pi)$ is a decreasing function of π , the number of fixed agents in the community in our model. Of course, $f(\pi)$ can simply be interpreted as an index of social interaction. However, it is useful to gain some intuition about what these results actually mean. One highly intuitive number to discuss is $1/\pi$, which in either model is the expected distance between two fixed agents. This distance is the expected size of an unbroken line of interconnected agents, or the expected size of a clique of interacting agents. We find it useful to think about the clique sizes implied by the empirical results. Viewed in this way, our empirical results can be seen as asking "how big must social groupings be to justify the observed cross-city variance?"

Formally, under the second model $f(\pi)=(2-\pi)/\pi$ or $1/\pi=(1+f(\pi))/2$. Using this formula, we can then translate our $f(\pi)$ estimates into estimates about clique size. In Table 2A, the values for $f(\pi)$ tell us that the average clique size for serious crimes in general ranges from 657 to 470 (for cross-city data).⁵⁴ The average clique size for robbery, burglary and assault are 192, 163 and 122 respectively for our largest value of λ . The average clique size for murder is estimated at five. From our most basic estimates, social interactions seem to flow over large groups for lesser crimes, but for murder and rape social interactions exist only within a unit the size of a modest family. In Table 3, if we assume that there are 6.4 crimes per criminal and $\lambda=0.04$, the clique size for serious crimes is 65. For auto theft in Table 3 the average clique size is still over 102.⁵⁵

In Table 4, when we have controlled for city level characteristics and estimated λ , estimated clique sizes for larceny and auto theft are over 190. Estimated clique sizes for robbery, assault and burglary are approximately 100.

⁵⁴Of course, since the serious crimes estimate assumes perfect cross-crime interactions and the other estimates assume no cross-crime interactions, the results are non-comparable.

⁵⁵ We view our auto theft results as conditional support for the multiple equilibrium model of automobile insurance advanced by Smith and Wright (1993).

Estimated clique size for arson is 17, rape is 8 and murder is 2. Using New York city data (again controlling for precinct level characteristics but assuming that $\lambda=0$), estimated clique size for serious crimes generally is 162, larceny is 75, auto theft is 51, robbery is 27, burglary is 30, assault is 6, murder and rape are one. Since the clique estimates for New York city data are approximately one-half to one third of the clique estimates that allow interactions to flow across districts, a rough estimate is that 40% of a clique lies within a single district and 60% of relevant social interactions occur with individuals outside the district.⁵⁶

Finally, our estimates from the 1986 data, which are, we believe, extreme lower bounds on the power of social interactions, show that the minimum clique size for assault, robbery and burglary lies between 25 and 35 actors. Clique size for larceny auto thefts are higher (over 50, although measured with a great deal more error). Our lower bound estimate of clique size for murder is one and rape is two, suggesting that there may be no significant social interactions in those crimes.

Geographic Clustering and the Age of Criminals

As mentioned previously, the observed social interactions could be measuring both local influences on choices about crime and criminals' moving into the same area. A full analysis of endogenous criminal migration is beyond our data and the scope of this paper,⁵⁷ but we believe that the age of criminals provides us with a natural method of examining the relative importance of migration and local influence. Adult criminals will have had more time to migrate and are more likely to have chosen the location they are inhabiting. Young agents (particularly those 18 and under) have little ability to choose their location. If the observed social interactions are the result of criminals' migrating, then we would expect the crimes committed by older criminals to display higher rates of social interactions.⁵⁸

⁵⁶In fact, this calculation is rough and formal analysis requires a more fully specified model of inter-district and intra-district social interactions.

⁵⁷We believe that micro data enabling us to look at the criminal behavior of migrants would enable us to separate these effects.

⁵⁸If the migration effect worked primarily through non-criminals leaving high crime areas, we would not expect to see these effects as strongly.

Alternatively, if social influence is most important in the young (perhaps because the young have less other information to contrast with peer influence), the social influence model predicts that younger crimes display more social interactions.

Figure 8 connects age with social interactions across the eight crimes we observe. These Figures regress the $f(\pi)$ estimates from Table 4 on the percentage of individuals arrested for each crime who are 18 and under in Figure 5.⁵⁹ The correlation between social interaction and youthfulness is strongly positive. While, we have only eight observations and other factors could easily be driving this correlation, the results suggest that the interactions we observe are more likely to be the result of local influence and not the result of geographic clustering of criminals.

Table 5 -- The Level of Interactions and City Characteristics

In the previous tables, we have assumed that $f(\pi)$ is invariant across cities. In fact, there are many reasons that we might believe the level of interactions changes across cities. Urban characteristics drive the mean level of crime rates, but they also may determine the extent to which agents interact and the extent to which patterns of crime move across the city unit. In this table, we extend the basic estimation used in equation (4.10). Both λ and $f(\pi)$ are estimated in a cross-city regression. However, now we no longer estimate $f(\pi)$ as a constant but allow it to change across city characteristics.

In particular, we allow three city characteristics to influence the degree of interactions: (1) percent nonwhite, (2) percent high school graduate and (3) percent female headed household. The race variable was included because a large, economically disadvantaged, minority may facilitate the social interactions that can cause crime. Education and the presence of a nuclear family may lessen the extent that neighbors provide valuable information about crime. Education provides an alternative source of information about the returns from criminal actions. Parents also may send messages that interrupt the messages from one's neighbors. We assume that $f(\pi)$ is a

⁵⁹The results are unchanged if we look at the share of arrestees who are 24 and under.

positive, linear function of urban characteristics (denoted $X'\beta$). Thus, we estimate:

$$(p_j - \hat{p}_j)^2 N_j = \hat{p}_j^2 (1 - \hat{p}_j)^2 (N_j - \text{Max}[0, X_j' \beta]) \Psi(\lambda, \hat{p}_j) + \text{Max}[0, X_j' \beta] \hat{p}_j (1 - \hat{p}_j) + \mu_j \quad (5.1)$$

where the X 's are the city-level characteristics that determine the overall level of interaction and again the \hat{p}_j 's represent the crime rates predicted in the first stage regression.

Our basic regression for 1985 found that race basically failed to influence the level of interactions except for murder and larceny, where bigger racial minorities created lower levels of interactions.⁶⁰ The level of education lowered the degree of interaction for murder and robbery, but raised the degree of interaction for larceny. The percent of female headed households raised the levels of social interaction for murder, serious crimes generally, burglary (at the 10% level) and larceny. The female-headed household variable had the expected sign and suggests that parental influence provides an alternative information source to peer effects. Perturbations of these basic models did not change any results significantly; the basic estimates of λ and the average $f(\pi)$ were robust to changes in the functional form and choice of explanatory variables.

Since there is a wide degree of heterogeneity across samples and across time, it is impossible to claim a victory for this methodology. It does seem true that a decrease in two-parent households increases criminal interactions in 1985 for larceny, auto theft and serious crimes. These results suggest that the presence of parents may limit the social interactions that create waves of petty crime or auto theft. However, this evidence is preliminary at best.

Tables 6A and 6B -- Multiple Equilibria Models

In the modeling section, we advanced the claim that local interactive models (of the sort we present) can better explain the data than a more traditional

⁶⁰Our larceny numbers differ significantly in magnitude from our other figures. These results seem to be correct and not the result of difficulty with our algorithm, but we are still somewhat less accepting of their accuracy.

multiple equilibria model such as Sah (1989) or Murphy, Shleifer and Vishny (1993). In a sense this is unsurprising, since the general interactive models have one extra degree of freedom and allow for results similar to those of the more traditional multiple equilibria models as special cases. The ultimate question for the utility of multiple equilibria models is whether they can explain a significant portion of the unexplained cross-city variance in crime rates. Tables 6a and 6b presents empirical evidence on this question.

Table 6A reestimates the variance of cross-city crime rates, but instead of assuming that these crime rates come from a single distribution, this Table assumes that the crime rates come from a mixture of normal distribution. These normal distributions are assumed to have the same variance (except for one row of the table) but they are allowed to have different means. This formulation is meant to capture the idea that each city's level of crime can come from one of several distributions. The goal is to see whether, once we've allowed urban crime rates to come from two (or more distributions), whether the remaining cross-city variance in crime rates is compatible with a model in which criminal decisions are independent across individuals.

Our methodology is based on the Dempster, Laird and Rubin (1977) EM (Estimation and Maximization) algorithm. Basically we estimate the distributions to maximize the likelihood function. Our primary finding is that allowing for multiple means does lower the estimate of cross-city variance, but not enough to eliminate the need for further explanations of the high levels of variance. When we allow for two normal distributions, we do find that the data can be split into two distributions, with 95.6% of the data belonging to the low crime distribution and 4.4% of the data coming from the high crime distribution. The mean crime rate for the low crime cities is .074 crimes per capita. The mean crime rate for the high crime distribution is .133 crimes per capita. The estimated variance for each of the two normals is 55.8 (rather than 96.9 for the one normal distribution).

In a way these results are supportive of a multiple equilibria model. Allowing two equilibria eliminates more than 40% of the cross-city variance to be explained. Unfortunately, the $f(\pi)$ estimated from the remaining variance is still over 750; there is still more than 750 times too much variance

to be compatible with an independent decision model. When we allow even more equilibria, the residual variance falls further but even with five distinct distributions, the remaining variance is still over 350 times the variance predicted by the independent model. With seven distributions, the remaining variance is 175 times the variance of the independent model.

We also estimated the two equilibria model allowing variances (as well as means) to differ across the two distributions being estimated. The results for this estimation are also in Table 6B. Now the data gets grouped into one high mean, high variance distribution and one low mean, low variance distribution. However, even the variance of the low mean, low variance, distribution is incompatible with independent decision making.

Overall, we believe that allowing multiple equilibria is not an effective method of explaining the high amount of cross-city variance in observed crime rates. It is true that in the extreme we could eliminate all of the cross-city variance by assuming a sufficiently large number of distributions that urban crime rates are drawn from, but our data suggests that even a model with five equilibria has a remaining variance 350 times the variance that we would see in the case where decisions are independent. The intuition of the multiple equilibria models, that we can explain high cross-city variance with interactions, is surely right, but their actual structure seems incompatible with the data.

Tables 7A and 7B -- Results on the Form of the Interaction

The previous results, and indeed the model, have been relatively vague about the form of the interaction across individuals. This section finally addresses possible forms of the positive interactions in crime. In general, we will examine four possible forms of interaction: (1) arrest rates, (2) returns to legal activities, (3) schooling and (4) parental stability. The theories behind all of these effects are fairly straightforward. The arrest rate theory is Sah's theory and it says that if my neighbor engages in crime, it makes it harder for the police to arrest me, so I will engage in more crime. The returns to legal activities theory says that if my neighbor steals from businesses, then it makes it harder for me to find a job in my neighborhood, so I will be more likely to

become a criminal. The schooling theory says that criminals make it harder to acquire education, because gangs make schools less productive, and with less schooling I am more likely to become a criminal. The single-parent family theory says that criminals are more likely to avoid family responsibility and that children raised without parents are more likely to become criminals.

If any of these four mechanisms lies behind the high variance of crime rates across cities, then we would not expect to observe high crime rates once we have controlled for those characteristics. However, as we showed in Table Four, controlling for all urban characteristics still left a large amount of unexplained urban crime. Still, one interpretation of our results is that the measures of these variables are actual noisy measures of the true variables and that if we had more accurate measures of the actual variable, the unexplained variance would disappear. So, we will re-examine these variables and when they actually are significantly related to observed crime rates, we will ask how much measurement error there would have to be for these mechanisms to actually explain the cross-city variance.

For serious crimes, and for the urban characteristics of (1) arrest rates, (2) unemployment rates, (3) schooling and (4) female-headed households, we have determined both the unconditional and conditional correlations between these characteristics and serious crimes (for 1970, 1985 and NYC precinct data). The conditional correlations are found by regressing both the crime rates and the urban characteristics on a vector of characteristics (population, population growth, four regional dummies and percent non-white).

The first potential interaction mechanism, arrest rates, is positively correlated with crime across NYC precincts, which perhaps occurs because police concentrate their resources more in high crime rate areas.⁶¹ Across cities in 1985 the correlation is -8% which means that arrest rates can explain less than

⁶¹In fact, if communities with higher crime rates hire more police officers and end up having higher arrest rates (as a positive correlation between crime and arrest rates indicates), it means that there is negative interaction term across individuals (which would eliminate cross-city variance completely not exacerbate that variance).

1% of the observed variation in crime rates. The strongest correlation between arrest rates and crime rates is in 1970, but this correlation is only -18% which still means that arrests are explaining less than 4% of crimes. These results are supported by the partial correlations, which reduce the explanatory power of the 1970 data. We must conclude that arrests play only a marginal role in explaining cross-city crime variance.

We obtained not only arrest data, but also conviction data across cities and again there is no correlation between the likelihood of being convicted and rates of crime. The conviction data shows only a 5% correlation (for our 15 cities in this sample) between conviction rates and crime rates. While these results are based on a small sample indeed, they do suggest that congestion in the courts is also not a candidate mechanism for positive social interaction.

Schooling is also strongly correlated with crime across cities. The correlation of high school graduation rates with serious crime rates is -25% in 1970 and -18% in 1986. Oddly across precincts in New York the correlation is positive. When we look at correlations that condition on demographics (city population, population growth and percent non-white) and regional dummies the correlation between schooling and crime falls to -9% in 1970 and -5% in 1985. These weak correlations make us doubt that schooling is the mechanism driving crime rates.

Unemployment rates are more correlated with crime rates. The range of correlation is between 20% (across NYC districts) and 33% across cities in 1970 in the United States for unconditional correlations. For conditional correlations, the 1985 correlation is 9% and the NYC precinct correlation is negative. Only for the 1970 data do we see a conditional correlation remain sizable at 20%.

Finally, the correlation between female headed households and serious crimes is 59% in 1970, 42% in 1985 and 22% across NYC precincts. The corresponding conditional correlation coefficients are 33%, 32% and -10% (across NYC precincts). While the conditional correlations are sizable and there is no doubt that family structure is correlated with criminal behavior,

the correlations are still far from explaining the bulk of the variance of crime rates.

Since both single-headed households and unemployment display sizable correlations, we will now ask how noisy our estimates of these variables would have to be for these variables to actually be the mechanisms driving criminal interactions. The noisiness could be that everything from keypunch error, to transitory (irrelevant) changes in the economic characteristics, to the fact that our variable is only a proxy for the correct variable. If a multiple equilibria model held, it would mean that:

$$P_j = \alpha + \beta Z_j + v_j, \text{ where } \text{Var}(v_j / \sqrt{N_j}) = p(1-p). \quad (5.2)$$

Equation (5.2) means that once we have controlled for Z_j (our true measure of the social interaction mechanism) the remaining variance is on the order of $p(1-p)$ (where p is, as earlier, the population average crime rate) which would be the variance if social interactions did not occur. Since $\text{Var}(v_j / \sqrt{N_j}) = p(1-p)$, $\text{Var}(v_j)$ is on the order of $1/N_j$, which we know is small and which we will ignore (as we did earlier).

We observe $Z_j + \xi_j$, not Z_j , where ξ_j is some independent measurement error.⁶² For each of our measures of social interaction, we are interested in knowing how close our measure is to the true aggregate interaction. Simple algebra, and using that $\text{Var}(v_j) = 0$, tells us that:

$$\text{Var}(\xi_j) / \text{Var}(Z_j) = (1 - R^2) / R^2 \quad (5.3)$$

where R^2 is the square of the correlation coefficient between Z_j and P_j .⁶³

The conditional correlation between female headed households and crime rates is 33% in 1985 ($R^2 = .109$). For this social variable to explain the cross-city variance, the ratio of noise to signal, $\text{Var}(\xi_j) / \text{Var}(Z_j)$, must be 8. For unemployment in 1970, the correlation coefficient is 9% and the ratio must be

⁶² This methodology is similar to that used in Murphy and Topel (1990).

⁶³ In fact equation (5.3) is not in any way a new result but simply a convenient way of thinking about the correlation coefficient.

over 100. These two variables tell us that if either model of aggregate interactions explains the bulk of crime rate variance, then our measures of economic conditions or social problems must be extremely poorly correlated with the relevant measures of these variables. However, this preliminary evidence does suggest that if a simple mechanism exists it is much more related to the social variable, female-headed households, than the more straightforward labor market variable, unemployment.

There is a second requirement, beyond eliminating the large cross-city variance, for a mechanism to be empirically palatable. This mechanism must both cause and be caused by crime. In all of these interactive models, the causality must be dual. While, in principle this interaction could be tested using instrumental variables, we have so far been unable to find convincing instruments with sufficient explanatory power for crime to test this implication adequately.

7. Conclusion

This paper has reexamined the extreme cross-city, and cross-precinct crime variance and found that it is essentially incompatible with a model where agents' decisions in a metropolis are independent. Our two models of social interactions provide a framework for understanding the observed variance of cross-city crime rates. More importantly, these models provide us with a natural index of social interactions. This index comes from the presence of "fixed" agents, or agents that do not themselves respond to their neighbor's propensity towards crime.

This model of fixed agents should not be taken literally. Instead, estimates of the number of fixed agents should be taken as a measure of the degree of interaction. Lesser or greater interaction could come from larger or smaller social groups and the number of fixed agents in the models directly determine the average size of a social group. More broadly, the amount of interaction could come from the ease of communication across individuals. A fixed agent can just be interpreted as an agent who does not see what his neighbor is doing.

We applied this index of social interactions to criminal behavior in the United States across cities and across precincts in New York. Even allowing for a wide diversity in underlying characteristics across cities (even more than is shown by observable characteristics), we found a large amount of social interaction in criminal behavior. The cross-city variance is just too high to be rationalized as the outcome of independent decisions to engage in crime -- there must be covariance across agents.

Our index showed that there is a wide range in the degree of interactions across crimes, but that across data samples, the rough level of interactions stayed constant for each crime. The estimates for average social group size ranged from 1 to 5 for murder. Similarly low levels of social group size were found for rape and arson. The average social group sizes for auto theft and larceny were over 200 in most of our estimations. For robbery, assault and burglary, estimated clique sizes was approximately 100.

Across crimes, we found that younger crimes have more social interactions. Across cities, we found the higher levels of social interactions (for serious crimes generally, for petty larceny and for auto theft) were found for cities with more female headed households. While these results are highly preliminary, we interpret them to mean that the average social interactions among criminals are higher when there are not intact family units. The presence of strong families interferes with the transmission of criminal choices across individuals.

We looked for mechanisms through which these social interactions occurred and found little conclusive evidence for any particular mechanism. Undoubtedly the forms that social interactions can take are varied and may include information, crowding of legal activities, and destruction of prevailing social structures. The determination of the forms of these social interactions must be left for future research.

References

- Archer, D. and R. Gartner (1984) *Violence and Crime in Cross-National Perspective*. New Haven: Yale University Press.
- Becker, G. S. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76: 169-217.
- Benabou, R. (1994) "Education, Income Distribution and Growth: The Local Connection," NBER working paper # 4798.
- Blumstein, A. and J. Cohen (1979) "Estimation of Individual Crime Rates from Arrest Records," *Journal of Criminal Law and Sociology* 70: 561-585.
- Bransom, M. and D. Griffeath (1979) "Renormalizing the three dimensional voter model," *Annals of Probability* 7: 418-432.
- Case, A. and L. Katz (1991) "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," NBER working paper.
- Chaikin, J. (1978) "Rand Prison Inmate Survey," Rand Working Paper WN-10107-DOJ.
- Crane, J. (1991) "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology* 96: 1226-1260.
- Dempster, A. P., Laird, N., and D. B. Rubin (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm," *The Journal of the Royal Statistical Society (Series B)* 39: 1-38.
- Ehrlich, I. (1975) "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *American Economic Review* 65: 397-417.
- Feller, W. (1968) *An Introduction to Probability Theory and Its Applications*. New York: John Wiley.
- Freeman, R. (1991) "Crime and the Employment of Disadvantaged Youths," NBER Working Paper # 3875.
- Glaeser, E., L. (1992) "Two Essays on Information and Labor Markets," University of Chicago Dissertation.
- Jacobs, J. (1961) *The Death and Life of Great American Cities*. New York: Random House.
- Kindermann, R. and J. L. Snell (1980) "On the relationship between markov random fields and social networks," *Journal of Mathematical Sociology* 7.
- Lane, R. (1979) *Violent Death in the City: Suicide, Accident and Murder in Nineteenth Century Philadelphia*. Cambridge: Harvard University Press.
- Levitt, S. (1994) "The Effect of Police on Crime," mimeographed.
- Murphy, K.M., A. Shleifer and R. Vishny (1993) "Why is Rent Seeking So Costly to Growth?" *American Economic Review* 83: 409-414.
- Murphy, K.M. and R. Topel (1990) "Efficiency Wages Reconsidered: Theory and Evidence," in Weiss and Fishelson Eds. *Advances in the Theory and Measurement of Unemployment*. London: Macmillan.
- Quetelet, A. (1835) *Sur L'Homme et le Developpement De Ses Facultes*. Paris: Bachelier.
- Sah, R. (1991) "Social Osmosis and Patterns of Crime," *Journal of Political Economy* 99: 1272-1295.

- Scheinkman, J. A. and M. Woodford (1994) "Self-organized Criticality and Economic Fluctuations," *American Economic Review* 84: 417-421.
- White, H. (1984) *Asymptotic Theory for Econometricians*. New York: Academic Press.
- Wilson, J. Q. and R. Herrnstein (1980) *Crime and Human Nature*. New York: Simon and Schuster.

Appendix: Variable Definitions

	<u>Definition</u>	<u>Source</u>
<i>Serious crime</i>	The sum of Part I offenses defined under the FBI's Uniform Crime Reporting program. Part I offenses include murder, rape, robbery, robbery, assault, burglary, larceny, and auto theft. The NYC precinct data exclude those Part I crimes that are not felonies, for example larcenies <\$1000. Hence NYC data are not fully comparable to cross city data. All NYC precinct data are for 1993.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Murder</i>	UCR definition: The willful and non-negligent killing of one human being by another. Excludes manslaughter by negligence and traffic fatalities.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Rape</i>	The carnal knowledge of a female forcibly and against her will. Excludes statutory rape.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Robbery</i>	The taking or attempt to take anything of value from the care, custody, or control of a person by force or threat of force.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Assault</i>	An unlawful attack by one person upon another for the purpose of inflicting severe or aggravated bodily injury.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Burglary</i>	The unlawful entry of a structure to commit a felony or a theft.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Larceny</i>	The unlawful taking, carrying, leading, or riding away of the property of another.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Auto theft</i>	The theft or attempted theft of a motor vehicle.	FBI Uniform Crime Reports; NYC Police Dept.
<i>Arson</i>	Any willful or malicious burning or attempt to burn, with or without attempt to defraud, a dwelling house, public building, motor vehicle or aircraft, personal property of another.	FBI Uniform Crime Reports
<i>Population, all samples</i>	City population as defined by Bureau of the Census	CPS; 1970 Census of Pop; 1980 Census of Pop;
<i>Regional dummies</i>	Uses regions (North, South, Central, West) as defined by Bureau of the Census	

	<u>Definition</u>	<u>Source</u>
<i>Persons over age 25 with 12 or more years school</i>		Census of Population
<i>Persons over age 25 with 4 or more years college</i>		Census of Population
<i>Unemployment rate</i>	Civilian labor force unemployment rate. Standard Labor Department definition	Bureau of Labor Statistics
<i>Percent households with female head</i>	Female householder. No spouse present.	Bureau of the Census
<i>Percent persons below poverty level</i>		Bureau of the Census
<i>Percent owner occupied housing</i>	Owner occupied housing units / occupied housing units	Bureau of the Census
<i>Property taxes per capita</i>	Total property taxes collected during fiscal year / population	Bureau of the Census
<i>Police per capita</i>	Full time city law enforcement officers / population	FBI
<i>Percent pop. non-white</i>		Bureau of the Census

Figure 1: $F(\pi)$ Values in the Two Models

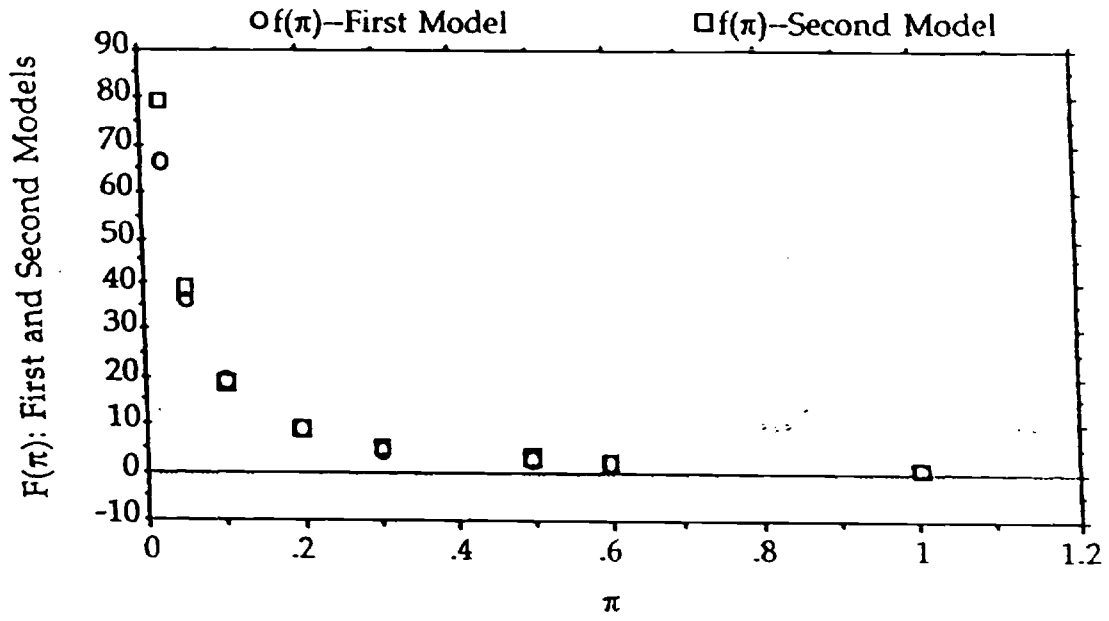


Figure 2: Serious Crimes Per Capita, 1985 U.S. Data

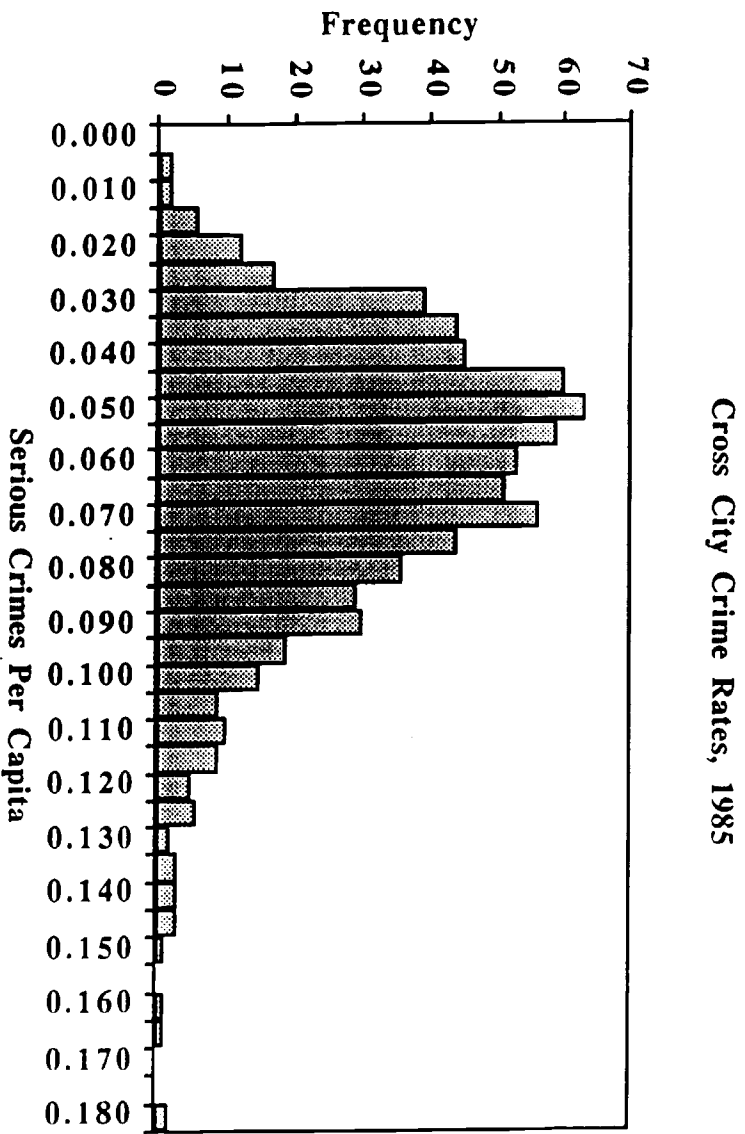
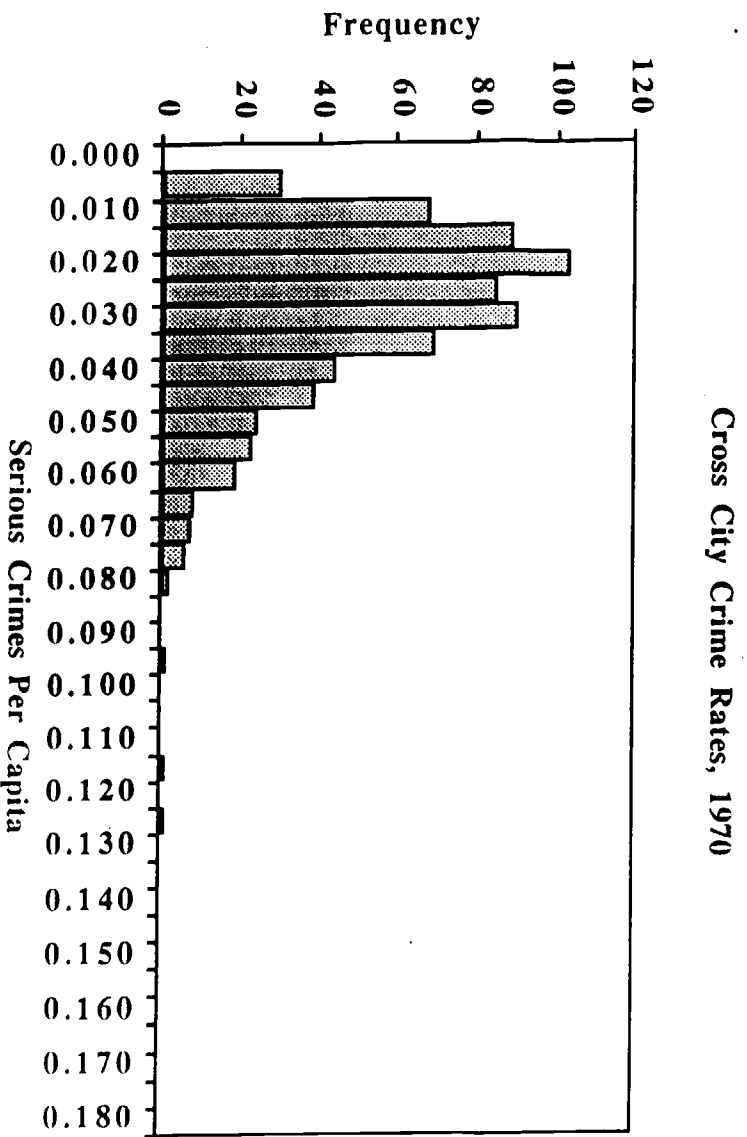
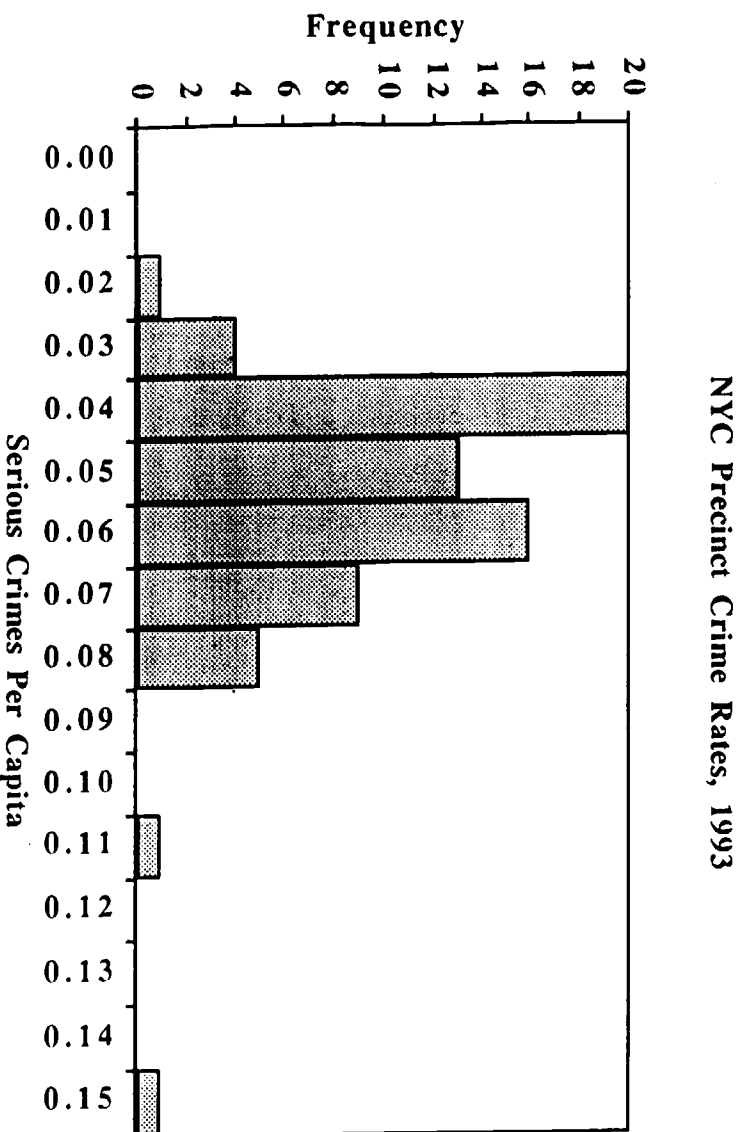


Figure 3: Serious Crimes Per Capita, 1970 U.S. Data



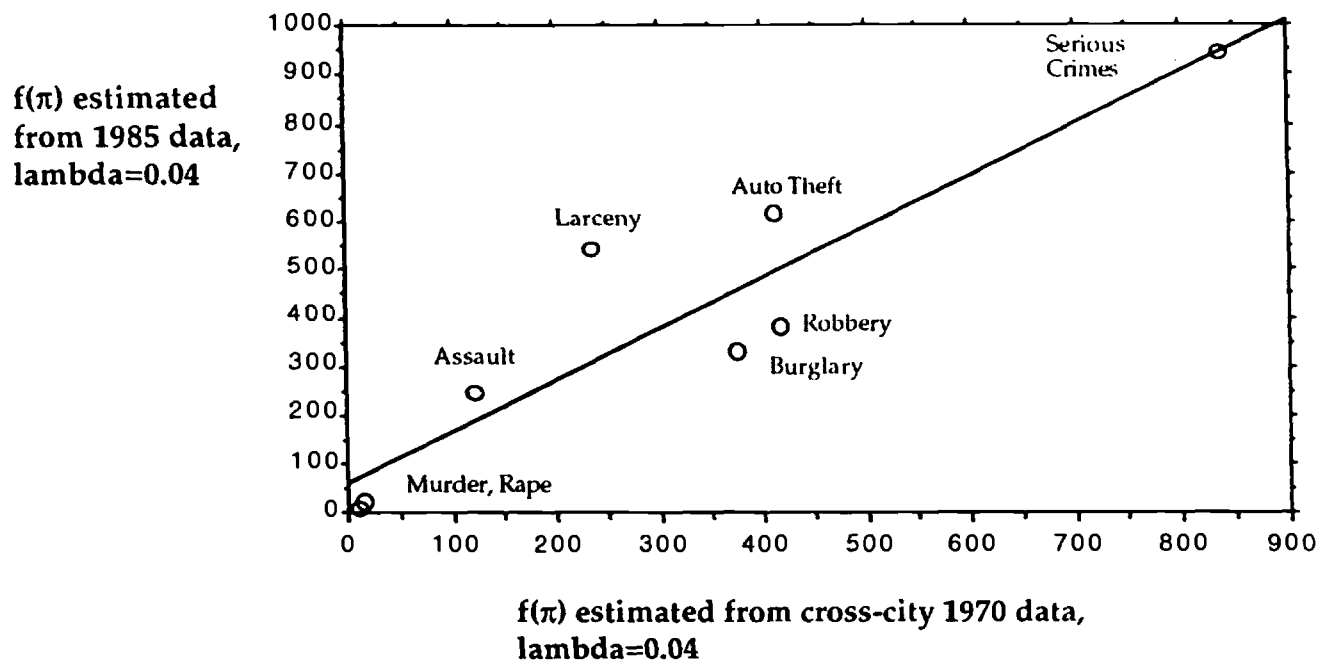
Notes: Sample defined as U.S. cities > 25,000 people. FBI data. See Appendix for sources and definition of serious crimes per capita.

Figure 4: Serious Crimes Per Capita, 1993 New York Data



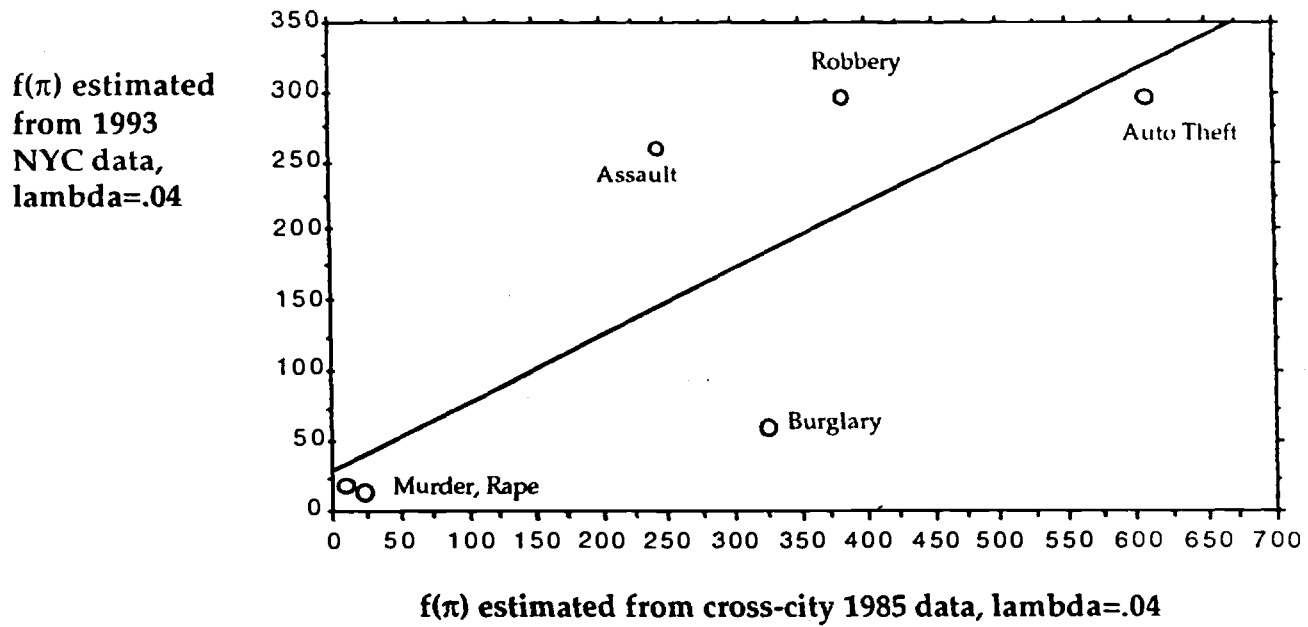
Notes: NY Police Department data. See Appendix for data sources and definitions.

Figure 5: $f(\pi)$ Estimates from Cross City Data, 1985



The data for this figure come from Table 2A.
Crime data ultimately comes from the FBI Uniform Crime Reports.

Figure 6: $f(\pi)$ Estimates from NYC and 1985 Cross-City Data

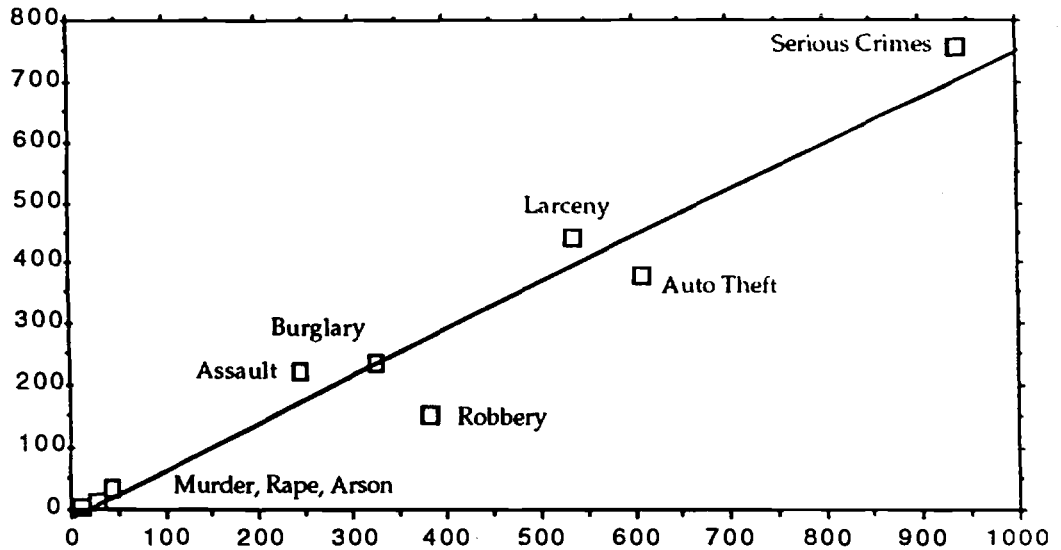


$$y = .483x + 29.354, R\text{-squared: } .62 \\ (.189)$$

$f(\pi)$ estimates are from Table 2A.
Crime data ultimately comes from NYC precinct data and the
FBI uniform crime reports.

Figure 7: $f(\pi)$ Estimates, when lambda is estimated and when lambda is fixed at 0.04.

$f(\pi)$ estimated from 1985 cross-city data, lambda is estimated, from Table 4.

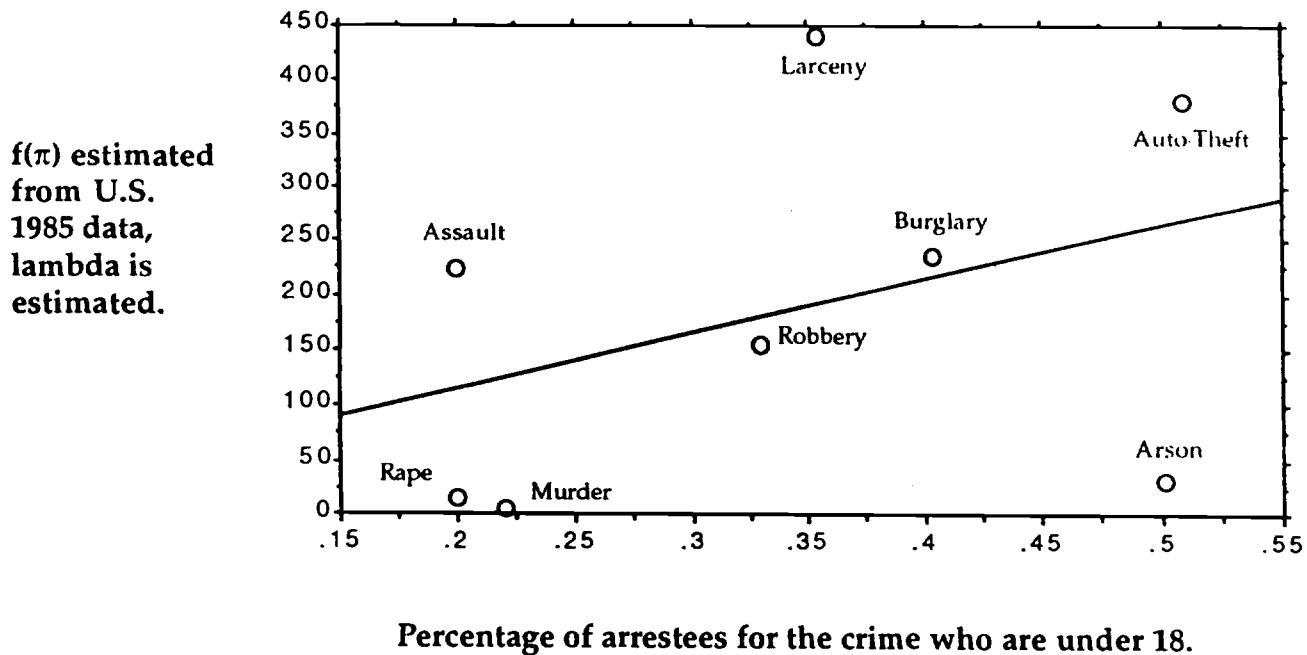


$f(\pi)$ estimated from cross-city 1985 data, lambda=.04, from Table 2A.

$$y = .761x - 14.507, R\text{-squared: } .942 \\ (.071)$$

The data for this figure come from Table 2A and Table 4 respectively. Crime data ultimately comes from the FBI Uniform Crime Reports.

Figure 8: $f(\pi)$ Estimates and the Age of Criminals



$$y = 497.758x + 16.95, R\text{-squared: } .144 \\ (495)$$

The $f(\pi)$ estimates come from Table 4.
Crime data comes from the FBI Uniform Crime Reports.

Table 1A: Five Highest and Five Lowest Crime Cities In the U.S.

City	Pop. 1986	Serious Crimes Per Cap. 1985	Serious Crimes Per Cap. 1970	% Fem. Heads House. 1980	Unemp. 1986	% HS Grads. 1980
Atlantic City, NJ	35,980	.384	.118	.407	.089	.481
West Palm Beach, FL	68,570	.181	.047	.219	.052	.657
Portland, OR	387,870	.169	.061	.178	.088	.758
Highland Park, MI	25,620	.165	.098	.453	.163	.566
Miami, FL	373,940	.151	.071	.237	.083	.500
Shelton, CT	34,840	.019	.014	.090	.039	.736
Melrose, MA	28,790	.018	.016	.134	.028	.832
Marlborough, MA	31,180	.015	.020	.148	.024	.756
Garfield, NJ	26,350	.012	.008	.160	.070	.449
Ridgewood Village, NJ	24,850	.008	.010	.100	.025	.908

Notes: Sample defined as U.S. cities > 25,000 people. Atlantic City and West Palm Beach have high crime rates in part because the city population (the denominator) does not capture the fact that so many people visit these cities each year.

Table 1B: Means and Standard Deviations, Cross City Data 1985

DATA SERIES	N	Mean	Min	Max	Standard Deviation
Serious crimes per capita	658	.0673	.0061	.3838	.0292
Murders per capita	658	7.597E-5	0	7.795E-4	8.472E-5
Rapes per capita	658	4.433E-4	0	.0024	3.507E-4
Robberies per capita	658	.0022	0	.0195	.0026
Assaults per capita	658	.0034	0	.0235	.0032
Burglaries per capita	658	.0156	0	.0573	.0080
Larcenies per capita	658	.0390	0	.2968	.0187
Auto thefts per capita	658	.0053	0	.0391	.0048
Arsons per capita	628	5.248E-4	0	.0053	5.105E-4
1986 population	658	128598	24610	7262750	362340
Pop. growth rate 1970-1986	658	13.5%	-31.5%	388%	36.6%
North (regional dummy)	658	.2067	0	1	.4052
South (regional dummy)	658	.2796	0	1	.4492
Central (regional dummy)	658	.2690	0	1	.4438
Persons over age 25 with 12 or more years school, 1980	658	68.4%	35.3%	96.4%	11.3%
Persons over age 25 with 4 or more years college, 1980	658	17.9%	4.7%	68.7%	9.5%
Unemployment rate, 1986	658	7.0%	0.7%	21.9%	3.1%
Percent households with female head, 1980	658	17.4%	6.2%	45.3%	6.0%
Percent persons below poverty level, 1979	658	12.3%	1.4%	36.4%	6.0%
Percent owner occupied housing	658	57.3%	13.4%	92.0%	12.4%
Property taxes per capita, 1985	658	170.29	1.48	1823.01	180.16
Police per capita, 1985	658	.0018	.0008	.0113	.0007
Percent pop. non-white, 1980	658	18.5%	0.6%	93.5%	16.7%

Note: Means and standard deviations are intentionally shown as unweighted and hence are not directly comparable to figures shown in tables 2-7.

Table 1C: Means and Standard Deviations, Cross City Data 1970

DATA SERIES	N	Mean	Min	Max	Standard Deviation
Serious crimes per capita	617	.0322	.0053	.1286	.0167
Murders per capita	617	6.789E-5	0	4.869E-4	7.500E-5
Rapes per capita	617	1.779E-4	0	.0015	1.692E-4
Robberies per capita	617	.0014	1.995E-5	.0165	.0019
Assaults per capita	617	.0016	2.712E-5	.0128	.0016
Burglaries per capita	617	.0123	.0012	.0438	.0070
Larcenies per capita	617	.0116	.0004	.0431	.0058
Auto thefts per capita	617	.0052	.0003	.0473	.0046
1970 population	617	124895	25131	7894851	391567
Pop. growth rate 1970	617	35.0%	-33.0%	2027%	121%
North (regional dummy)	617	.2091	0	1	.4070
South (regional dummy)	617	.2804	0	1	.4496
Central (regional dummy)	617	.2626	0	1	.4404
Persons over age 25 with 12 or more years school, 1970	617	56.4%	25.0%	90.0%	12.0%
Persons over age 25 with 4 or more years college, 1970	617	12.1%	3.0%	46.0%	7.3%
Unemployment rate, 1970	617	4.1%	1.0%	12.0%	1.6%
Percent households with female head, 1970	617	11.6%	4.0%	27.0%	4.0%
Percent persons below poverty level, 1970	617	11.9%	2.1%	45.4%	6.3%
Percent owner occupied housing	617	58.8%	14.0%	93.0%	12.8%
Property taxes per capita, 1970	617	69.71	4.00	1178.00	79.42
Police per capita, 1970	617	.0016	.0006	.0067	.0006
Percent pop. non-white, 1970	617	11.6%	0.1%	73.9%	13.6%

Note: Means and standard deviations are intentionally shown as unweighted and hence are not directly comparable to figures shown in tables 2-7.

Table 1D: Means and Standard Deviations, New York City Data

DATA SERIES	N	Mean	Min	Max	Standard Deviation
Serious crimes per capita	70	.0605	.0216	.2059	.0236
Murders per capita	70	.0003	0	.0011	.0003
Rapes per capita	70	.0004	2.897E-5	.0012	.0003
Robberies per capita	70	.0129	.0009	.0326	.0068
Assaults per capita	70	.0064	.0010	.0201	.0044
Burglaries per capita	70	.0139	.0049	.0278	.0045
Larcenies per capita	70	.0119	.0031	.1175	.0153
Auto thefts per capita	70	.0145	.0040	.0371	.0076
1990 population	70	100316	32978	221763	46046
Persons over age 25 with 12 or more years school, 1990	70	65.7%	36.3%	95.2%	14.6%
Persons over age 25 with 4 or more years college, 1990	70	21.8%	3.4%	66.1%	16.9%
Unemployment rate, 1990	70	10.6%	3.5%	23.3%	5.0%
Percent households with female head, 1990	70	31.5%	6.8%	63.4%	15.5%
Percent persons below poverty level, 1990	70	21.7%	3.8%	53.2%	13.7%
Percent pop. non-white, 1990	70	59.6%	9.7%	99.4%	31.1%

Note: Means and standard deviations are intentionally shown as unweighted and hence are not directly comparable to figures shown in tables 2-7.

Table 1E: Means and Standard Deviations, Cross City Data, 1986

DATA SERIES	N	Mean	Min	Max	Standard Deviation
Serious crimes per capita	631	.0728	.0088	.3854	.0330
Murders per capita	631	8.627E-5	0	8.575E-4	9.348E-5
Rapes per capita	631	.0005	0	.0031	.0004
Robberies per capita	631	.0026	0	.0225	.0029
Assaults per capita	631	.0042	0	.0308	.0039
Burglaries per capita	631	.0169	.0017	.0623	.0086
Larcenies per capita	631	.0424	.0061	.2978	.0215
Auto thefts per capita	631	.0061	.0006	.0408	.0055
Arsons per capita	578	.0006	0	.0048	.0005

Note: Means and standard deviations are intentionally shown as unweighted and hence are not directly comparable to figures shown in tables 2-7.

Table 2A: Variance and Decomposition

DATA SERIES	P(1-P)	Variance (1)	f(π) $\lambda=0$	f(π) $\lambda=.004$	f(π) $\lambda=.04$	λ^* f(π)=1	Variance (2)
Serious Crimes 1985, N=658	0.073	95.91	1313.8	1276.7	941.1	0.14	.042
1970, N=617	0.042	43.91	1045.5	1024.7	837.1	0.20	.017
NYC ¹ , N=70	0.053	30.48	575.1	553.9	363.2	0.11	.021
Murder 1985, N=658	0.0001	0.001	10.0	9.9	9.5	0.70	0
1970, N=617	0.0001	0.001	10.0	10.0	9.5	0.72	0
NYC, N=70	0.0003	0.006	20.0	19.9	18.8	0.63	0
Rape 1985, N=658	0.0006	0.016	26.7	26.4	23.6	0.33	0
1970, N=617	0.0003	0.005	16.7	16.5	15.2	0.42	0
NYC, N=70	0.0004	0.007	17.5	17.3	15.9	0.41	0
Robbery 1985, N=658	0.0047	1.92	408.5	406.1	384.4	0.67	.0003
1970, N=617	0.0037	1.61	435.1	433.3	416.7	0.94	.0004
NYC, N=70	0.011	3.75	340.9	336.5	296.9	0.31	.0020
Assault 1985, N=658	0.0048	1.29	268.8	266.3	244.1	0.43	.0005
1970, N=617	0.0025	0.337	134.8	133.6	122.3	0.43	.0001
NYC, N=70	0.0054	1.52	281.5	279.3	259.9	0.52	.0008
Burglary 1985, N=658	0.019	8.057	424.1	414.3	326.6	0.17	.003
1970, N=617	0.016	7.25	453.1	445.2	373.4	0.23	.003
NYC, N=70	0.013	1.46	112.3	107.1	60.2	0.09	.008
Larceny 1985, N=658	0.04	29.75	743.8	723.3	538.9	0.14	.015
1970, N=617	0.012	3.53	294.2	288.2	234.3	0.20	.001
NYC ² , N=70	0.01	11.23	1123.0	1119.0	1083.3	1.12	.008

Table 2A: Variance and Decomposition, Continued

Data Series	P(1-P)	Sample Variance (1)	f(π) $\lambda=0$	f(π) $\lambda=.004$	f(π) $\lambda=.04$	λ^* f(π)=1	Sample Variance (2)
Auto theft 1985, N=658	0.008	5.21	651.3	647.2	610.3	0.63	.001
1970, N=617	0.008	3.63	453.8	449.8	413.9	0.45	.001
NYC, N=70	0.015	5.35	356.7	350.7	296.7	0.24	.003
Arson 1985, N=628	0.00074	0.033	44.6	44.2	40.8	0.46	.0001

Table 2B: Variance and Decomposition, Other Variables

Data Series	P(1-P)	Sample Variance (1)	f(π) $\lambda=0$	f(π) $\lambda=.004$	f(π) $\lambda=.04$	λ^* f(π)=1	Sample Variance (2)
# of Females 1970 N=617	0.25	24.24	97.0	-28.0	-1163.6	0.00	.011
High School Graduates 1980, N=658	0.23	534	2321.7	2205.5	1149.2	0.08	.26
High School Graduates 1970, N=617	0.22	1472.2	6691.8	6587.7	5642.4	0.24	.68

The first column gives number of crimes per capita in each sample minus that number per capita squared. The second and seventh columns gives the variance of normalized crime levels (described formally below) across cities or precincts.

P=population probability of crime, N_i=population size,

Second Column:

$$\text{Sample variance (1)} = \frac{\sum_{i=1}^{\# \text{ cities}} [(\text{crime city}_i - P_{us} * N_i) * \sqrt{N_i}]^2}{\# \text{ of cities}}$$

Seventh Column:

$$\text{Sample variance (2)} = \frac{\sum_{i=1}^{\# \text{ cities}} [(\text{crime city}_i - P_{us} * N_i) * \sqrt[5/6]{N_i}]^2}{\# \text{ of cities}}$$

The third through fifth columns give the amount of local interaction (denoted f(π)) implied by these means and variances for given λ levels, where λ is described in the text. The sixth column gives the level of λ for which there is no local interaction.

Table 3: Variance and Decomposition Controlling For Crimes Per Criminal

DATA SERIES	P(1-P)	Variance (1)	f(π) $\lambda=0$	f(π) $\lambda=.004$	f(π) $\lambda=.04$	λ^* f(π)=1	Variance (2)
Serious Crimes 1985, N=658 Crime per=3	0.026	10.7	411.5	398.2	278.1	0.12	.005
1985, N=658 Crime per=6.4	0.012	2.3	191.7	185.5	130.0	0.12	.001
1985, N=658 Crime per=10	0.0079	0.96	121.5	117.5	80.9	0.12	.0004
Robbery 1985, N=658 Crime per=3.3	0.0014	0.16	114.3	113.6	107.1	0.63	.00005
Assault 1985, N=658 Crime per=1.7	0.0028	0.45	160.7	159.3	146.3	0.44	.0002
Burglary 1985, N=658 Crime per=5.4	0.0035	0.28	80.0	78.2	62.0	0.18	.0001
Larceny 1985, N=658 Crime per=10.44	0.004	0.27	67.5	65.4	46.9	0.13	.0001
Auto Theft 1985, N=658 Crime per=2.8	0.003	0.66	220.0	218.5	204.6	0.57	.0002

Crimes-per-criminal data is derived from two sources. For all serious crimes, the value of 6.4 is derived from a self-reported measure described in the Rand Prison Inmate Survey (Chaiken, 1978). For individual crimes, data on crimes-per-criminal stems from Blumstein and Cohen's (1979) study of arrest records.

Table 4: Variance and Decomposition Controlling For City Specific Characteristics

Data Series	P(1-P)	Sample Variance	Variance $(p - \hat{p})\sqrt{N}$	$f(\hat{\pi})$ $\lambda = 0$	$\hat{\lambda}$	$f(\hat{\pi})$
Serious crimes 1985, N=658	0.073	95.91	49.6	932.3 (111.4)	.013 (.003)	754.6 (118.2)
1970 N=617	0.042	43.91	13.7	522.5 (41.3)	.004 (.001)	475.1 (42.5)
NYC N=70	0.053	30.48	15.4	324 (109)	1	
1986 N=631	0.078	117	11.93	159.7 (53.7)	.0003 (.0015)	155.0 (58.5)
Murder 1985, N=658	0.0001	.001	.0003	5.6 (.4)	.012 (.002)	4.49 (.46)
1970 N=617	0.0001	.001	0.0002	4.0 (0.3)	0 ²	4.0 (0.3)
NYC N=70	0.0003	.006	0.0005	1.34 (.24)	1	
1986 N=631	0.00015	.0017	0.0002	2.62 (.19)	.0005 (.0009)	2.58 (.21)
Rape 1985, N=658	0.0006	.016	0.0006	17.1 (1.0)	.011 (.003)	14.8 (1.2)
1970 N=617	0.0003	.005	0.0021	15.4 (1.8)	.005 (.006)	14.6 (2.0)
NYC N=70	0.0004	.007	0.0008	1.16 (.33)	1	
1986 N=631	0.0006	.017	0.0024	4.8 (0.4)	.002 (.001)	4.4 (0.4)
Robbery 1985, N=658	0.0047	1.92	0.28	155.0 (13.2)	0 ²	155.0 (13.2)
1970 N=617	0.0037	1.61	.153	111.0 (12.5)	0 ²	111.0 (12.5)
NYC N=70	0.011	3.75	.713	53.6 (13.5)	1	
1986 N=631	0.005	2.00	.141	52.2 (7.8)	0 ²	52.2 (7.8)
Assault 1985, N=658	0.0048	1.29	.67	244.6 (17.0)	.014 (.004)	224.0 (18.0)
1970 N=617	0.0025	.337	.0001	115.4 (10.1)	.002 (.003)	113.1 (10.6)
NYC N=70	0.0054	1.52	0.112	13.4 (2.7)	1	
1986 N=631	0.0056	1.66	0.286	57.4 (11.1)	0 ²	58.3 (12.1)

Table 4: Variance and Decomposition Controlling For City Specific Characteristics, continued

Data Series	P(1-P)	Sample Variance	Variance $(p - \hat{p})\sqrt{N}$	$f(\hat{\pi})$ $\lambda = 0$	$\hat{\lambda}$	$f(\hat{\pi})$
Burglary 1985, N=658	0.019	8.057	4.35	341.0 (28.1)	.027 (.003)	236 (30)
1970 N=617	0.016	7.25	3.03	297.6 (25.1)	.009 (.002)	257 (26)
NYC N=70	0.013	1.46	.83	61.1 (11.4)	1	
1986 N=631	0.02	9.61	1.02	69.2 (5.7)	.0011 (.0007)	63.9 (6.5)
Larceny 1985, N=658	0.04	29.75	18.23	590 (61)	.024 (.005)	441 (67)
1970 N=617	0.012	3.53	2.19	196.8 (17.7)	.005 (.002)	186.4 (18.2)
NYC N=70	0.01	11.23	3.4	150.0 (39.2)	1	
1986 N=631	0.042	38.05	7.83	146.5 (89.1)	.0005 (.0065)	143.3 (100.5)
Auto Theft 1985, N=658	0.008	5.21	1.82	509.7 (39.5)	.024 (.004)	382.2 (42.7)
1970 N=617	0.008	3.63	1.19	340.3 (31.1)	0 ²	340.3 (31.1)
NYC N=70	0.015	5.35	1.50	102.1 (17.7)	1	
1986 N=631	0.009	5.84	.72	122.7 (9.2)	.001 (.0009)	118.0 (10.2)
Arson 1985, N=628	0.00074	.034	0.017	33.1 (3.6)	0 ²	33.1 (3.6)
1986 N=578	0.00077	.035	0.006	11.7 (0.7)	.0001 (.0009)	11.7 (.8)

The first and second columns are the same as those in Table 2. The third column represents the variance adjusted for observable city tendencies towards crime:

$$\text{Variance} = \frac{\sum_{i=1}^{\# \text{ cities}} [(\text{crime city}_i - PE_i * N_i) * \sqrt{N_i}]^2}{\# \text{ of cities (second column)}}$$

where PE_i =predicted (based on OLS regressions controlling for current population, past population growth, regional dummies, education rates, unemployment rates, female headed households rates, poverty levels, owner occupied housing rates, property taxes, police per capita, and percent non-white and for 1986, 1985 crime rates). The fourth column is the $f(\pi)$ implied using that variance. The fifth and sixth columns are the estimated levels of λ and $f(\pi)$.

1. Estimation of λ for the New York City data sample is impossible because there is insufficient cross-precinct population variance (the precincts were chosen to be of the same size).

**Table 5: Variance and Decomposition,
Allowing $f(\pi(z))$ to Vary with City Characteristics**

DATA SERIES	λ	Average $f(\pi)$	Determinants of $f(\pi)$				R ²
			Intercept	Percent nonwhite	Percent HS Grad	Percent Female HH	
Serious Crimes 1985, N=658	.010 (.003)	895	-1723 (1192)	216 (1048)	1010 (1291)	9311 (3251)	.09
1970, N=617	.004 (.001)	488	270 (366)	172 (368)	-133 (450)	1814 (1583)	.12
NYC, N=70	.1 ¹	182	-3843 (3498)	-1961 (1752)	7673 (4272)	-3183 (4079)	.17
Murder 1985, N=658	.011 (.002)	4.1	14.6 (4.9)	-8.2 (3.2)	-17.4 (6.0)	17.0 (10.9)	.21
Rape 1985, N=658	.012 (.003)	14.7	18.9 (11.6)	-5.3 (8.7)	-6.3 (13.5)	7.5 (27.2)	.20
Robbery 1985, N=658	0 ²	176	563 (160)	-9.3 (106)	-385 (182)	-626 (344)	.13
Assault 1985, N=658	.012 (.004)	214	408 (186)	132 (162)	-290 (215)	-186 (479)	.16
Burglary 1985, N=658	.024 (.004)	267	-78 (297)	223 (256)	35 (329)	1283 (797)	.17
Larceny 1985, N=658	.030 (.004)	95	-60,000 (10,000)	-12,000 (3,000)	36,000 (7,000)	138,000 (19,000)	.36
Auto theft 1985, N=658	.023 (.004)	361	399 (442)	151 (203)	-181 (508)	199 (969)	.19
Arson 1985, N=658	0 ²	36	23 (39)	-2 (26)	26 (45)	-22 (97)	.02

The averages were found by averaging the predicted values of $f(\pi)$ over cities, weighting by city population. The dependence of $f(\pi)$ on city characteristics assumed a linear form where $f(\pi)$ is equal to $X'B$ for X being a vector that includes a constant (to estimate the intercept) and the city features. $f(\pi)$ is constrained to be >0 .

1. Estimating λ is impossible for New York because of insufficient cross-precinct population variation.

Table 6A: Reduction In Variance Implied By Fitting Crime Rates to a Mutiple Equilibria Model

# means	σ^2	μ_a	μ_b	μ_c	μ_d	μ_e	μ_f	μ_g
1	95.9	.0797 <i>1.0</i>						
2	55.8	.074 <i>.956</i>	.133 <i>.044</i>					
3	35.1	.061 <i>.703</i>	.087 <i>.254</i>	.136 <i>.044</i>				
4	34.1	.055 <i>.224</i>	.066 <i>.513</i>	.088 <i>.218</i>	.136 <i>.044</i>			
5	28.5	.054 <i>.348</i>	.067 <i>.391</i>	.088 <i>.213</i>	.133 <i>.044</i>	.216 <i>.044</i>		
6	18.4	.052 <i>.450</i>	.071 <i>.30</i>	.087 <i>.16</i>	.11 <i>.05</i>	.14 <i>.03</i>	.41 <i>.0015</i>	
7	14.3	.051 <i>.460</i>	.072 <i>.29</i>	.087 <i>.15</i>	.11 <i>.06</i>	.13 <i>.03</i>	.16 <i>.01</i>	.41 <i>.0015</i>

Notes: N=658 Data used are serious crimes in 1985 divided by the square root of population. Uses EM algorithm (Dempster, Laird, Rubin, 1977) to separate the data into 2-7 normals with 2-7 means, but same variance. Mixing probabilities (i.e. the probability of city_i being in a given equilibrium) are shown in italics.

Table 6B: Estimation of Two Separate Normals (Two Means, Two Variances)

# means	σ_a^2	σ_b^2	μ_a	μ_b
2	24.21	194	.064 <i>.78</i>	.096 <i>.22</i>

Table 7A: Raw Correlations Between Crime Rates and Potential Sources of Social Interaction

	Arrest Rate	HS Grad. Rate	Unemp. Rate	% Female Heads House	Conviction Rate
Serious Crime Rate 1985	-.08	-.19	.26	.42	not available
Serious Crime Rate 1970	-.18	-.23	.33	.59	.05
Serious Crime Rate NYC Precincts 1993	.25	.03	.20	.22	not available

Notes: N=659 for 1985 cross city data. N=617 for 1970 data. N=70 for NYC precinct data. N=15 for conviction rate correlation. Variable definitions and sources are in appendix. Each crime variable is correlated with corresponding interaction variable (e.g. cross-city crime rate 1970 is correlated with 1970 arrest rate, NYC crime rate correlated with NYC arrest rate, etc.)

Table 7B: Partial Correlations Between Crime Rates and Potential Sources of Social Interaction-- After Controlling For Covariates

	Arrest Rate	HS Grad. Rate	Unemp. Rate	% Female Heads House
Serious Crime Rate 1985	-.05	-.05	.09	.33
Serious Crime Rate 1970	-.11	-.05	.20	.32
Serious Crime Rate NYC Precincts 1993	.03	.32	-.12	-.10

Notes: Uses the residuals from OLS regressions on population, population growth rate, 4 regional dummies, and percent non-white.