

Structural Estimation of a Becker-Ehrlich Equilibrium Model of Crime: Allocating Police Across Cities to Reduce Crime

CHAO FU

University of Wisconsin

and

KENNETH I. WOLPIN

Rice University and University of Pennsylvania

First version received September 2015; Editorial decision October 2017; Accepted November 2017 (Eds.)

We develop a model of crime in which the number of police, the crime rate, the arrest rate, the employment rate, and the wage rate are joint outcomes of a subgame perfect Nash equilibrium. The local government chooses the size of its police force and citizens choose among work, home, and crime alternatives. We estimate the model using metropolitan statistical area (MSA)-level data. We use the estimated model to examine the effects on crime of targeted federal transfers to local governments to increase police. We find that knowledge about unobserved MSA-specific attributes is critical for the optimal allocation of police across MSA's.

Key words: Crime, Multiple equilibria, Estimation, Efficient police allocation.

JEL Codes: J2, K42

1. INTRODUCTION

The modern literature on the economics of crime, originating with [Becker \(1968\)](#), recognized that the crime rate is the equilibrium outcome of the joint determination of the supply of crime, resulting from the uncoordinated decisions of citizens about the supply of their labour to legal and/or illegal income-generating activities, and the demand for crime, determined by a government policymaker who decides on the level of resources to commit to preventing crime.¹ Much of the empirical literature, beginning with [Ehrlich \(1973\)](#), has used aggregate data based on either cross-sectional

1. For a recent updated version of this framework, see [Ehrlich \(2010\)](#).

or time-series variation.² Coupled with the supply/demand theoretical framework, the use of aggregate data led naturally to the adoption of a simultaneous equations econometric structure.³ The estimating equations in that system consisted of a supply of offences function representing the decision rule of potential criminals, an apprehension production function and the policymaker's decision rule governing the level of resources devoted to apprehension (and punishment).

The econometric structure was meant to approximate the solution to the equilibrium model. As such, its parameters are combinations of those of the underlying behavioural structure, that is, of the preference function of potential offenders (the citizenry), the apprehension production function, the objective function of the policymaker and the distribution of the unobservables that enter those functions. An alternative approach, pursued in this article, is to estimate the behavioural structure by specifying and solving a parametric model of agent decision-making. This approach relaxes two potentially important aspects of the approximation in the traditional approach. First, it explicitly aggregates the individual decisions about whether to engage in illegal activities over the citizen population, which naturally leads to nonlinearities and location-specific effects. The existence of location-specific effects, as derived from the model, implies that allowing for random coefficients would be a more appropriate approximation in the traditional regression-based approach. This observation is particularly important in the context of instrumental variables (IVs) (see Angrist and Imbens, 1994; Heckman and Vytlačil, 1998) as is a common procedure used in estimating the deterrent effect of apprehension with aggregate data.

Secondly, this approach accounts for the existence of multiple equilibria that is inherent in the structure. Multiple equilibria arise because the probability that a criminal is apprehended is assumed to depend on the number of crimes committed, which creates a positive externality for engaging in crime. The existence of multiple equilibria confounds the interpretation of regressions in the traditional approach. To see how, note that because each city faces different fundamentals, the equilibrium sets differ across cities. Such sets may differ both in the number of equilibria in the set and by the equilibrium outcomes (the crime rate and the apprehension rate) within the set. In order that the regression coefficient on, for example, the apprehension rate reflect what would happen to the crime rate within a location if the apprehension rate were to exogenously change, the goal of estimation, it is necessary that the equilibrium selected across locations that differ in their apprehension rates is in some sense the same. Examples of equilibrium selection rules that satisfy a sameness criterion include choosing the equilibrium with the lowest crime rate or choosing the Pareto-dominant equilibrium (if one always exists). If different cities select different equilibria, then the regression coefficient would reflect that change in equilibria.

The model that we estimate specifies the optimization problem of potential criminals and of the policymaker in a city. Each city has a continuum of individuals of different types, unobservable to the researcher, in terms of their legal sector human capital and in terms of their preferences for staying at home and for committing crimes. Types are correlated with observable characteristics, age, gender, education, and race. Besides the distributions of their citizens' characteristics, cities also differ in their production technology, in their apprehension efficiency and in the opportunity cost of the resources devoted to law enforcement. Each citizen chooses whether to be a criminal, work in the legal sector, or stay at home. City output is produced with the human capital of citizens who choose to work in the legal sector. Criminals meet victims randomly and receive a

2. For a recent review, see Tauchen (2010). Durlauf *et al.* (2010) point out some limitations in using aggregate crime regressions to analyse crime determinants and evaluate counterfactual policies.

3. This contrasts with the empirical literature on labour supply that took a decidedly partial equilibrium approach given the availability of micro-level data (*e.g.*, Heckman, 1971). However, recent studies have used longitudinal data to estimate the supply of crime function, for example, Lochner and Moretti (2004) and Imai and Krishna (2004). For an early specification of dynamic models of criminal choice, see Flinn (1986).

fraction of the victim's legal sector income. For a given city, the arrest rate varies with the crime rate and the size of the police force. Given any size of the police force, it is feasible to solve for all possible equilibria because the model structure yields an ordering over citizens in their propensity to be criminals that is invariant across equilibria. Acting as a Stackelberg leader, the policymaker chooses the number of police to maximize the expected value of an objective function that includes the number of police (negatively), the crime rate (negatively) and the apprehension rate (positively), where the expectation is taken over the distribution of equilibria that we estimate. The solution to the model yields equilibrium values of the number of police, the employment rate, the crime rate, the apprehension rate, and the competitively determined rental price of human capital.

The estimation of the model is by simulation. At any given set of parameter values and set of location-specific unobservable characteristics, the model can be solved for the equilibrium outcomes for each location and for all of the equilibria. Doing multiple simulations for the same parameter values by randomly drawing from the distribution of unobservables and averaging over the simulations within each location provides statistics for the equilibrium outcomes that can be matched with data. The data are from two sources: the Uniform Crime Reports (UCRs) of the FBI and the Current Population Survey (CPS). The CPS provides demographics, employment, and wage data, while the UCR provides data on crimes, arrests and the number of police. We apply the model to property crimes. We estimate the model using data on the 238 metropolitan statistical area (MSA)'s that can be matched between the two data sets for the year 2008. We use data for the year 2003 to conduct an out-of-sample validation. On the whole, we find the out-of-sample fit to be reasonably good.

With our estimated model, we examine the factors underlying the different crime rates across MSA's: the observable and unobservable characteristics of the MSA's and their citizens, and the selection of different equilibria. We develop a method for distinguishing the extent to which crime rate differences are due to differences in observable characteristics versus differences in the equilibrium that is selected based on a metric we propose for comparing equilibria across cities. We illustrate the decomposition by comparing pairs of MSA's. We find that there is heterogeneity across pairs in the relative importance of equilibrium selection. For example, controlling for unobservables, the crime rate in Atlanta would fall by 15% if the equilibrium selection had been the same as in Philadelphia, resulting in a crime rate essentially equal to that in Philadelphia. On the other hand, the difference in crime rates between Houston and Philadelphia is hardly driven at all by the different equilibria that are selected. In addition, conditioning on the equilibrium selection being the same across MSA's, we find that observable MSA characteristics also affect crime rates; for example, crime rates are higher in MSA's with a younger and less educated population. These findings are qualitatively similar to that found in the "reduced form" literature which does not account for multiple equilibria.

We then conduct two sets of counterfactual policy experiments that are not feasible within the conventional simultaneous equations estimation framework using cross-sectional data. These experiments are motivated by federal programmes that provide subsidies to local governments for increasing the number of police. For example, the Community Oriented Police Services (COPS) programme, initiated by the Clinton administration in 1994, aimed at a nationwide increase in the number of police of 20%.⁴ In all of the experiments, we assume that the federal government can perfectly monitor the use of the resources to any locality, ensuring that the intended increase in the size of the police force is realized. In the first set of experiments, we explore two scenarios. In the first, unlike the COPS programme, the planner (federal government) uniformly increases

4. See [Zhao *et al.* \(2002\)](#) and [Evans and Owens \(2007\)](#) for an analysis of that programme.

the size of the police force by 20% for each MSA; the program leads to an 8.2% reduction in the national crime rate.

In the second scenario, as in the COPS programme, instead of a uniform transfer, the planner subsidizes newly hired police, where the number of new hires is chosen by each local government. First, we determine the subsidy rate such that the planner's total spending is the same as under the uniform transfer. At such a subsidy rate (48%), there is a 42% increase in the number of police and the national crime rate falls by 21%, or 2.6 times the decrease in the uniform policy. Secondly, we determine the subsidy rate such that the increase in police nationwide, hence the total cost (federal and local), is the same as in the uniform transfer case, *i.e.*, a 20% increase. The subsidy rate in that case is 27%, which leads to a 10.6 percent decrease in crime, as compared to 8.2% in the uniform transfer case. Allowing local governments to optimally choose the number of police, instead of a uniform transfer, leads to a greater reduction in crime. Finally, we determine the subsidy rate such that the total spending mimics the intended cost of the COPS programme. Such a subsidy rate (34%) leads to a 26% increase in the number of police, as compared to the 20% goal set by COPS, and reduces the crime rate by 13%, which would be the effect of a programme like COPS if the federal government were able to perfectly monitor the use of its grants.

Our second set of experiments explores the effects of various targeting schemes to allocate federally sponsored additional police across locations, given a fixed resource constraint. We illustrate the idea by focusing on pair-wise allocation problems, where the total additional resource to be allocated between a pair of MSA's is equivalent to 20% of their current total police force. To achieve the optimal targeting, we first use the model estimates, for each MSA, to determine the values of the MSA-level unobserved characteristics: the MSA's arrest efficiency, marginal cost of police, value of leisure and productivity. We explore five different allocation rules for each of three different planner objective functions. The first allocation rule assumes the planner has information on both the observable and unobservable characteristics of the MSA's and optimizes over the objective function. The second rule assumes the planner has information only on observables and optimizes over the objective function. The other three rules are not based on optimizing an objective function, but instead allocate police using either current crime rates, current arrest rates, or current GDP per capita. We study the cases for five pairs of MSA's. In the case of Philadelphia versus Phoenix, for example, the allocation with complete information, based on minimizing the overall crime rate, leads to a 16% decrease in the crime rate, while the crime reduction ranges from 5.6% (allocation based on crime rates) to 10.1% (allocation based on observables) across the other four allocation rules. Although the allocation with complete information always dominates the other rules, the relative effectiveness of the other four allocation rules is found to be case-dependent.

1.1. *Related literature*

Our article contributes to the literature that studies crime from an equilibrium perspective. Via different channels, various theoretical studies have demonstrated the existence of multiplicity of equilibria in models of crime. For example, [Fender \(1999\)](#) studies a model of crime where agents with different earnings abilities decide whether to work or to commit crimes. The arrest rate depends on the crime rate and the expenditure on law enforcement. They show that multiple equilibria may exist, and that for certain parameter values there may be multiple stable equilibrium levels of crime. [Conley and Wang \(2006\)](#) study an equilibrium model where agents, with heterogenous working abilities and tastes for crime, choose either to commit crimes, or invest in education and become workers. The arrest rate depends on the crime rate and the number of police. They establish that when individuals differ in more than one dimension, multiple interior equilibria with different positive crime rates may exist. Taking the arrest rate as an exogenous

parameter, [Burdett *et al.* \(2003, 2004\)](#) introduce crime into an otherwise classical random search equilibrium framework, where firms post wages and meet workers with an exogenous probability.⁵ Besides the random job offers they receive, workers (unemployed or employed) may also receive a criminal opportunity at random. Workers choose whether or not to accept the job offer in the case they receive one and whether or not to commit a crime in the case they receive a criminal opportunity. Multiple interior equilibria with a positive crime rate may arise due to matching externalities.⁶ Our article is the first to empirically implement a model of crime with multiple equilibria.

[Imrohoroglu *et al.* \(2004\)](#) develop and calibrate a dynamic (supply-side) equilibrium model to study the trend in crime between 1980 and 1996. Taking the arrest rate as exogenous, they study individuals' dynamic choices of whether or not to be a criminal after a stochastic period-specific employment status is realized. An equilibrium requires, among others, that the aggregate crime rate is consistent with individuals' choices. [Imrohoroglu *et al.* \(2000\)](#) embed a static equilibrium model of crime in a political economy framework. Individuals choose to specialize in either legitimate or criminal activities. The police force is funded by tax revenues from labour income, where the tax rate is determined via a majority-voting rule. The number of police is the sole determinant of the arrest rate, regardless of the crime rate. In [Fella and Gallipoli \(2014\)](#), heterogeneous agents live for three periods: a schooling period, work period, and retirement period. An agent decides on savings/consumption (with borrowing constraints) in each period, chooses an educational level in the first period and in the second period inelastically supplies labour and decides on crime involvement while facing an exogenous arrest rate per crime. This life cycle model is embedded in a general equilibrium framework, where the rental rates for human capital and physical capital are determined in competitive markets. The government administers a pay-as-you-go pension system, collects taxes to maintain the justice system (a fixed cost per convicted criminal) and to spend on transfers and wasteful public expenditure. Using their calibrated model, they find that, compared to lengthening sentences, subsidizing high school graduation (balanced via adjusting income tax rates) is much more effective in reducing crime and improving total welfare.

[Adda *et al.* \(2014\)](#) study the impact of a cannabis depenalization experiment within an equilibrium framework. The experiment depenalized cannabis purchase in the London borough of Lambeth. The model incorporates the decision by potential cannabis consumers as to their location of purchase, which depends on the allocation of police resources in Lambeth and all other London boroughs. The supply of other non-drug-related crimes is modelled similarly. Police resources in each borough are allocated between apprehending drug and non-drug offenders. The equilibrium rate at which crimes of each type is detected is determined by the interaction of the crime-specific supply of offenders and supply of police. The model is used to perform the counterfactual of depenalizing cannabis purchases city-wide. They find that cannabis consumption rises modestly in all boroughs. Moreover, non-drug crimes fall in all boroughs as police are reallocated to those crimes.

Our article complements these papers by incorporating the direct impact of the crime rate on the arrest rate within a general equilibrium framework and confronting the potential multiple equilibria problem that results. Given the complications arising from potential multiple equilibria, we do not consider the political economy aspect studied in [Imrohoroglu *et al.* \(2000\)](#) or the life

5. Examples of other papers that build crime in a search model framework include [Engelhardt *et al.* \(2008\)](#) and [Huang *et al.* \(2004\)](#).

6. Besides matching externalities and multi-dimensional heterogeneity among agents, multiple equilibria in models of crime may also arise from other sources, for example, geographical externalities (*e.g.* [Freeman *et al.*, 1996](#); [Verdier and Zenou, 2004](#)) and direct interpersonal spillovers (*e.g.* [Sah, 1991](#); [Glaeser *et al.*, 1996](#)).

cycle aspect studied in [Fella and Gallipoli \(2014\)](#). As with [Adda *et al.* \(2014\)](#), we focus on the allocation of police resources, though in our case in the context of a centralized policy.⁷ In order to feasibly capture the equilibrium aspect of the determination of crimes, our model necessarily makes some simplifications. For example, as discussed in the review by [Draca and Machin \(2015\)](#), one can allow for, though we do not, the choice between criminal and legal activities as an allocation problem with continuous hours, allow for different types of criminal specialization (*e.g.* shoplifting versus auto theft) and/or differential (monetary) payoffs from crime and allow for dynamics. Incorporating some of these aspects into an equilibrium setting like ours would comprise a substantial agenda for future research.

In the simultaneous equations framework, it is possible to derive a relationship between crime and police. Estimating that full structure would enable one to determine how an exogenous increase in police would affect the apprehension rate and how the apprehension rate affects criminal behaviour. There have been many criticisms of this methodology.⁸ The more recent literature that has sought to estimate the effect of police on crime has adopted a “search for instruments” approach. Using our model estimates to mimic that approach, we shed light on the interpretation of IV estimates in the literature. In particular, we show how IV estimates vary with the particular sample and with the available instrument.

The rest of the article is as follows. The next section describes the model. Section 3 describes the data and Section 4 our estimation strategy and results. Section 5 explains the difference across MSA’s in terms of their crime rates. Counterfactual experiments are presented in Section 6. The last section concludes the article. Some details and additional tables are in the Appendix.⁹

2. MODEL

There are J cities $j=1, \dots, J$, each with a government and a continuum of individual citizens. Cities are considered as closed economies.¹⁰ Each government acts as a Stackelberg leader by choosing the size of its police force.¹¹ Observing their government’s decision, individuals in each

7. There is also a related operations research literature that takes a game-theoretic approach to the police allocation decision problem. That literature is primarily concerned with developing real-time algorithms for solving the game-theoretic model for the optimal amount of police resources to apply to specific geographic areas within a designated location, for example, a city. Examples of that literature include [Zhang *et al.* \(2015\)](#), [Paruchuri *et al.* \(2008\)](#) and [Zipkin *et al.* \(2014\)](#). The models that are posited in this literature are well suited to the short-run police allocation decision they are designed to solve, but less well suited to the policies we study. For example, those models take the number of criminals and police as given whereas we posit those as choices.

8. See, for example, [Durlauf *et al.* \(2010\)](#).

9. Supplementary data and codes are available as [Supplementary materials](#) at Review of Economic Studies online.

10. The model abstracts from mobility between cities (MSA’s in our empirical analysis). We view the baseline outcomes in the data as representing the steady state. In that case, mobility is not an issue in estimation. However, policy counterfactuals would have to take account of mobility to the extent that cross-city mobility may be induced by the policy. It is likely that mobility would mute the effect of policies that make criminal activity more (or less) attractive in some cities than in others, although less so in the short-run and/or if policy changes are small and mobility costs are high. To the extent that mobility is consequential, our counterfactual policy implications are better interpreted as short-run equilibrium effects. In contrast, the importance of mobility considerations is highlighted in the growing literature on “hot spots”, which studies the distribution of criminal activities within, rather than between, cities. Allowing for mobility at the individual level and for strategic interactions among cities that would naturally result are beyond the scope of this article.

11. A game where the government and citizens move simultaneously would admit even a larger set of equilibria. In addition, it is reasonable to assume that the government moves first because government budgets are typically set annually.

city choose one of three mutually exclusive and exhaustive discrete options: work in the legal sector, work in the criminal sector, or remain at home.¹²

Each citizen is endowed with a human capital level (l), a taste for crime (η) and a value of “leisure” when at home (κ). The triplet (l, η, κ) defines an individual’s type, which is unobservable to the researcher. Each component of an individual’s type is assumed to be discrete with $l \in \{l_1, \dots, l_{N_l}\}$, $\eta \in \{\eta_1, \dots, \eta_{N_\eta}\}$ and $\kappa \in \{\kappa_1, \dots, \kappa_{N_\kappa}\}$. Therefore, there are $N = N_l \times N_\eta \times N_\kappa$ types of individuals. We let $n \in \{1, \dots, N\}$ be the index of a type defined as $(l_n, \eta_n, \kappa_n) \in \{l_1, \dots, l_{N_l}\} \times \{\eta_1, \dots, \eta_{N_\eta}\} \times \{\kappa_1, \dots, \kappa_{N_\kappa}\}$ and denote the proportion of individuals of type n in city j as p_{jn} . The type proportions are city-specific and related to observable characteristics (x): age, gender, race, education, and the number of young children in the household, all treated as discrete where the number of discrete values are 4, 4, 2, and 3. In total, there are 192 distinct demographic groups. The distribution of x , $G_j(x)$, is city-specific. We denote the discrete choices for a type n individual as $d_{n1} = 1$ if working in the legal sector (=0 otherwise), $d_{n2} = 1$ if at home (=0 otherwise) and $d_{n3} = 1$ if working in the criminal sector (=0 otherwise).

2.1. The legal sector ($d_{n1} = 1$)

Legal sector output in city j , Y_j , is produced using the aggregate stock of human capital of those citizens in city j who choose to work in that sector, L_j . The production technology is given by

$$Y_j = \tau_j L_j^\theta,$$

where $\theta \in (0, 1)$ is the elasticity of output with respect to aggregate human capital and τ_j is a city-specific Hicks-neutral technology factor drawn from the distribution $\tau_j \sim \ln N(-0.5\sigma_\tau^2, \sigma_\tau^2)$. Assuming a competitive labour market in each city, the rental rate for a unit of human capital is given by its marginal product,

$$r_j = \tau_j \theta L_j^{\theta-1}, \quad (1)$$

and earnings for an individual of type n residing in city j , y_{jn} , is the product of the rental price in city j and the individual’s level of human capital, that is, $y_{jn} = r_j l_n$.

2.2. The home sector ($d_{n2} = 1$)

The income of those who choose to be at home, denoted as b_j , is assumed to be equal to the human capital rental price in the city times the lowest level of human capital, l_1 . Because we do not distinguish between non-participation and unemployment, their income is intended to capture both unearned income and unemployment insurance. It is set to a low value, although one that varies with a city’s productivity via r_j , because those sources of income comprise, on average, a small proportion of total household income across all households within the population.

2.3. The criminal sector ($d_{n3} = 1$)

Each law-abiding citizen in city j , whether working or at home, faces an equal probability μ_j of being the victim of a crime. If victimized, an individual loses a fraction α of his income to the

12. The assumption that working in the legal and illegal sectors are mutually exclusive choices is required both by our data (see below) and as accommodation to computational feasibility. See, for example, Grogger (1998) for studies that treat activities in both sectors as a time-allocation problem.

criminal.¹³ A citizen who chooses to work in the criminal sector faces probability π_j of being arrested. The probability of an arrest (the arrest production function) depends positively on the size of the police force, s_j , and negatively on the crime (victimization) rate, μ_j . There is a city-specific unobserved component of technology, ϵ_j , which captures the unobservable factors that affect the efficiency of criminals in avoiding arrests, or equivalently police inefficiency.¹⁴ A city with a lower value of ϵ_j , that is, with higher police efficiency, has a higher arrest rate for a given number of police. The arrest technology function is given by

$$\pi_j = \Pi(s_j, \mu_j, \epsilon_j) = \exp\left(-\frac{\gamma(\mu_j \epsilon_j)^\rho}{s_j}\right), \quad (2)$$

where $\rho > 0$, $\epsilon_j \sim \ln N(-0.5\sigma_\epsilon^2, \sigma_\epsilon^2)$, and $\gamma > 0$ is a normalizing constant.¹⁵ The functional form ensures that $\Pi(\cdot) \in [0, 1]$. Notice that the arrest rate declines with the crime rate and that the rate of decline depends on whether $\rho \gtrless 1$. Note also that the parameterization of the degree of police (in)efficiency, which has mean one and is multiplicative with the crime rate, implies that a city with twice the crime rate and twice the police efficiency as another city will have the same arrest rate (given the same number of police).

2.4. The individual's decision problem

The decision of an individual in city j depends on his type, (l, η, κ) , as well as on city-level variables: the crime rate (μ_j), the human capital rental rate (r_j), the aggregate labour input (L_j), and the arrest rate (π_j). It will also depend on the expected utility of being a successful criminal (A_j), where the expectation is taken over the victim's income.¹⁶ Define the vector $\Omega_j(s_j) \equiv [\mu_j, r_j, L_j, \pi_j, A_j]$, which will vary with the number of police (s_j).¹⁷ Flow utility is assumed to be logarithmic in disposable income (consumption) and additive in the taste for crime and in the value of staying home. We also assume that criminals cannot target their victims and that they cannot steal from other criminals. Letting $d_n = [d_{n1}, d_{n2}, d_{n3}]$ be an individual of type n 's choice vector, the alternative-specific values for such an individual residing in city j is given by

$$V_{nj}(d_n | \Omega_j(s_j)) = \begin{cases} \mu_j \ln((1-\alpha)y_{jn}) + (1-\mu_j)\ln(y_{jn}) & \text{if } d_{n1} = 1, \\ \mu_j \ln((1-\alpha)b_j) + (1-\mu_j)\ln(b_j) + \kappa_n & \text{if } d_{n2} = 1, \\ \mu_j \ln(b_j) + (1-\mu_j)[\pi_j \ln(c) + (1-\pi_j)A_j] + \eta_n & \text{if } d_{n3} = 1. \end{cases} \quad (3)$$

The first (second) row in (3) shows the value if the individual chooses to work (stay home). With probability μ_j , the individual is victimized and consumes $(1-\alpha)y_{jn}$ if employed or $(1-\alpha)b_j$

13. We restrict attention only to property crimes.

14. Given that the only information available about crime is the aggregate crime rate, it is necessary to rule out heterogeneity of criminal abilities across individuals within a city; by including ϵ_j , we allow for such heterogeneity at the city level.

15. In the implementation, γ is further parameterized to be a deterministic function of the education distribution of the city. Specifically,

$$\gamma_j = b_0 + b_1 \ln(1 + P_j(\text{edu} = 1)),$$

where $b_0 > 0$, $b_1 \geq 0$ and $P_j(\text{edu} = 1)$ is the fraction of citizens in city j without a high school degree. Our counterfactual results are robust to a less restrictive specification where the marginal distributions of all demographics (education, age, race, gender, and number of children) affect γ_j , as shown in Appendix A4. The results we report are based on the more parsimonious specification.

16. The formulation for A_j is given below when a market equilibrium is defined.

17. To save on notation, we have suppressed the dependence of each component in the vector on s_j . In addition, we include both L_j and r_j in the individual's state space for completeness in describing the vector of outcomes although only one is necessary for the individual's choice problem.

if at home. With probability $(1 - \mu_j)$, the individual is not victimized and consumes his income. If he chooses to be at home, he also enjoys the value of staying home, κ_n . The third row shows the value if the individual chooses to be a criminal. With probability μ_j , a criminal fails to find a victim (criminals cannot be victims); in that case, we assume that the criminal has the same income as a law-abiding non-worker, b_j . With probability $(1 - \mu_j)$, a criminal meets a victim. In this case, with probability π_j , he is arrested and punished, consuming \underline{c} .¹⁸ With probability $(1 - \pi_j)$, he is not arrested, and has expected utility A_j . Engaging in crime also directly increases (or decreases) utility by the value η_n .

2.4.1. Optimal decisions. It can be shown that an individual in city j of type (l_n, η_n, κ_n) will engage in crime if and only if¹⁹

$$(1 - \mu_j)[\pi_j \ln(\underline{c}) + (1 - \pi_j)A_j] + \mu_j \ln\left(\frac{b_j}{1 - \alpha}\right) - \ln(r_j) > \max\{\ln(l_n), \ln(l_1) + \kappa_n\} - \eta_n. \quad (4)$$

If the individual does not choose to be a criminal, the individual will choose to work if only if

$$\ln(l_n) \geq \ln(l_1) + \kappa_n. \quad (5)$$

We denote the optimal decision of an individual by $d_n(\Omega_j(s_j))$.

From condition (4), it can be seen that an individual's propensity to engage in crime can be summarized by

$$T_n \equiv \max\{\ln(l_n), \ln(l_1) + \kappa_n\} - \eta_n. \quad (6)$$

We index these criminal propensities such that $T_n \leq T_{n+1}$, in which case the lower is n , the higher is one's criminal propensity. Thus, if T_n type chooses to be a criminal, all $T_{n'}$ types will do so for $n' < n$.²⁰

2.5. Market equilibrium

Definition 1. Given the size of the police force s_j , a market equilibrium in city j consists of a vector $\tilde{\Omega}_j(s_j) = [\mu_j, r_j, L_j, \pi_j, A_j]$, together with a set of optimal individual decision rules $\{d_n(\cdot)\}$ for $n = 1, \dots, N$ such that

(a) for all n , $d_n(\tilde{\Omega}_j(s_j))$ is an optimal decision for type n , i.e. conditions (4) and (5) hold;

18. We assume that regardless of how much they steal, the punishment is such that a criminal consumes \underline{c} if apprehended. This assumption is consistent with the fact that punishment is greater for robbery, which presumably has a higher payout, consumed by the criminal, than larceny.

19. We assume that an individual who is indifferent between legal and illegal activities will choose to be decent. This assumption is not without loss of generality given that types are discrete; however, our counterfactual experiment results are robust to this assumption.

20. The key assumptions that allow for the ranking structure are (1) Individuals within a city are homogeneous with respect to b_j , \underline{c} , A_j and non-employment value. (2) Individual differences in the expected utility from crime and home are captured by additively separable preference heterogeneity (η_n and κ_n), and utility is logarithmic in consumption.

(b) $\tilde{\Omega}_j(s_j)$ is consistent with individual choices where

$$\text{crime rate} : \mu_j = \sum_{n=1}^N p_{jn} d_{n3}(\tilde{\Omega}_j(s_j)), \quad (7)$$

$$\text{rental rate} : r_j = \tau_j \theta L_j^{\theta-1},$$

$$\text{aggregate labor} : L_j = \sum_{n=1}^N p_{jn} l_n d_{n1}(\tilde{\Omega}_j(s_j)),$$

$$\text{arrest rate} : \pi_j = \Pi(s_j, \mu_j, \epsilon_j),$$

crime utility | not apprehended :

$$A_j = \sum_{n=1}^N p_{jn} \frac{1 - d_{n3}(\tilde{\Omega}_j(s_j))}{1 - \mu_j} \ln \left(\alpha \left[\frac{y_{jn} d_{n1}(\tilde{\Omega}_j(s_j))}{+ b_j d_{n2}(\tilde{\Omega}_j(s_j))} \right] + b_j \right).$$

In the equation for A , with probability p_{jn} , a criminal meets a type- n citizen in city j , who can be a victim if and only if he is not a criminal ($1 - d_{n3}(\cdot) = 1$). Besides the government subsidy b_j , a get-away criminal also consumes an α fraction of the victim's income, which is y_{jn} if the victim works ($d_{n1}(\cdot) = 1$) and b_j if he stays home ($d_{n2}(\cdot) = 1$). Integrating over all types and dividing the integrated value by the probability that the criminal meets a victim ($1 - \mu_j$), one obtains the criminal's expected pecuniary utility conditional on successfully stealing from a victim (and not being arrested), *i.e.* A_j as in the value function (3).

As is clear from the arrest technology, a higher crime rate implies a lower arrest rate for a given size of the police force, thereby increasing the expected value of becoming a criminal. That is, there exists a positive externality among criminals. As a result, multiple $\Omega_j(s_j)$'s can be supported as market equilibria. However, as seen from (6), if T_n type chooses to be a criminal, all $T_{n'}$ types will do so for $n' < n$; and the ranking of T_n types is independent of equilibrium objects. Thus, all of the equilibria can be strictly ordered by their equilibrium crime rates, the latter being the total measure of types that choose to be criminals. Letting h_{jn} denote the measure of type T_n in city j and $H_{jn} = \sum_{n' \leq n} h_{jn'}$ the cumulative distribution of criminal propensities, then $\Omega_j^n(s_j)$ with $\mu_j^n = H_{jn}$ can be supported as an equilibrium if

$$T_n < (1 - \mu_j^n) \left[\pi_j^n \ln(\underline{c}) + (1 - \pi_j^n) A_j^n \right] + \mu_j^n \ln \left(\frac{b_j^n}{1 - \alpha} \right) - \ln(r_j^n) \leq T_{n+1}, \quad (8)$$

where the superscript n indexes the n^{th} potential equilibrium.

In equilibrium $\tilde{\Omega}_j^n(s_j)$, an individual will choose to be a criminal if and only if the individual's criminal propensity is ranked among the top n groups. The total number of equilibria, N^* , is no greater than the total number of T_n types and is bounded above by $N_\eta \times (N_k + N_l - 1)$.²¹ The fact that there are at most N^* equilibria and that they can be ordered, as given by (8), allows us to compute all of the market equilibria given s_j .²²

21. Note that N^* is parameter-dependent. See Appendix A3 for details about N^* and h_{jn} .

22. If individuals' criminal propensities cannot be ranked independent of equilibrium objects, there will be 2^N potential market equilibria; and it will be infeasible to compute all possible equilibria at every candidate parameter configuration during estimation.

2.6. *Government problem*

A government cares about the crime rate in the city and can affect the level of criminal activity by choosing the size of the police force. In addition, a government may also care directly about the arrest rate for political reasons; a government without the ability to catch criminals could be considered as ineffectual in combatting crime. Therefore, the government’s loss function will be positively related to the crime rate and, given its cost, to the size of the police force and negatively related to the arrest rate. The government is assumed to minimize its expected loss, where the expectation is taken over all possible market equilibria. Formally, the government in city j solves the following problem

$$\min_{s_j} \left\{ \sum_{n=1}^{N^*} q_j^n(s_j) \left(\omega_1 \exp(\mu_j^n) - \omega_2 \ln(\pi_j^n) + v_j s_j \right) \right\}, \tag{9}$$

where ω_1 and ω_2 are the weights governments put on the crime and arrest rates relative to the cost of the police force. $v_j \sim \ln N(-0.5\sigma_v^2, \sigma_v^2)$ is the city-specific marginal (opportunity) cost of the police force.²³ $q_j^n(s_j)$ is the probability that $\Omega_j^n(s_j)$ is realized as a market equilibrium, to be specified in Section 4. Although given any s_j there may be multiple market equilibria, the government optimal choice is generically unique.²⁴

2.7. *Subgame perfect Nash equilibrium*

Definition 2. A subgame perfect Nash equilibrium in city j is $\{s_j^*, d(\cdot), \{\tilde{\Omega}_j^n(\cdot)\}\}$ such that

- (a) Given any s_j , $(d(\cdot), \tilde{\Omega}_j^n(s_j))$ is a market equilibrium;
- (b) s_j^* solves the government’s problem.

3. DATA

We make use of data from two sources. One source is the CPS, which provides micro-level data on demographics (x), wages and employment. We focus on the population aged 16 to 64.²⁵ We aggregate individuals within a MSA, which is the counterpart of a “city” in our model. The distribution of x within an MSA is taken as the “city-specific” distribution of x in our model. We define an individual as employed if he/she works for more than 13 weeks during the year, and define the employment rate as the fraction of those employed among the age 16–64 population.²⁶ For those who are employed, we also use information on their annual earnings. We focus on MSA’s instead of cities because city identity is not available in the CPS and a reporting unit in the crime data may cover multiple cities.

The second data source is the UCR, which contains agency-level reports on crimes, arrests, number of police, and population size. As in previous studies (e.g. Imrohorglu *et al.*, 2000, 2004), we focus on property crimes, which include robbery, burglary, larceny-theft, and motor vehicle theft.²⁷ We made this choice because our model is more suitable to study crimes that

23. In addition to the direct cost of crime prevention, v_j captures the differential needs for resources across cities in areas other than crime prevention.

24. See Appendix A3 for the proof.

25. The age restriction is consistent with the literature, for example, Imrohorglu *et al.* (2000) (2000, 2004).

26. Our definition of employment follows the literature, e.g. Imrohorglu *et al.* (2000).

27. Note that our definition of property crimes is consistent with the literature but different from the UCR definition, which doesn’t include robbery.

TABLE 1
Summary statistics: x distribution

%	Education		Age		Race	Gender	Num. Young children	
	< HS	≥ BA	< 25	> 50	Black/Hispanic	Male	0	1
Mean	16.08	24.32	21.43	25.75	24.07	49.32	84.09	10.85
Std. Dev.	6.99	9.34	5.03	6.78	18.70	3.80	4.95	4.16

Number of Obs: 238 MSA's in the U.S.

are mainly driven by financial incentives. Each agency is supposed to report on a monthly basis. However, not all agencies have data for all months. It may be inappropriate to assume that crime patterns in the months reported are representative of that in the whole year, given for example, the well-known seasonality of crimes. Therefore, within each MSA, we aggregate only agencies for which data are available for all 12 months of the year. These agencies account for 70% of the total and 92% of the population covered in the UCR. As shown in Appendix Table A11, the population covered by agencies with non-missing reports is typically much greater than that covered by agencies with missing reports (average population size 23,400 versus 4,800). Although we cannot compare crime rates or arrest rates across these two groups of agencies (90% of the agencies with missing reports did not report crime statistics in any month), police-to-population ratios are similar. Both the population coverage and the representativeness of the police-to-population ratio among the agencies in our final sample suggest that our results may not be sensitive to this sample selection. For each MSA, we define the crime rate as the total number of actual crimes divided by the total 16–64 population covered by non-missing agencies, the arrest rate as the total number of arrests divided by the total number of actual crimes and the size of the police force as the total number of police divided by the total 16–64 population covered by non-missing agencies.

For both the CPS and the UCR, we estimate our model using data for the year 2008, the most recent year for which we have usable agency-level data. In 2008, the CPS data covers 265 MSA's and the UCR data covers 366 MSA's (including those with some non-missing agencies). We are able to match 245 MSA's between the two data sets. We exclude seven MSA's with zero arrests as extreme outliers, which results in a final sample consisting of 238 MSA's and 86,248 individuals living in these MSA's in the CPS sample. As shown in Appendix Table A12, the MSA's we use for estimation on average are larger than the ones we fail to match between the two data sets (average population of 588,410 versus 142,700). However, the two groups of MSA's have similar crime rates, arrest rates and police-to-population ratios. These facts suggest that the dropping the unmatched MSA's should not be problematic.

Table 1 summarizes the MSA-specific marginal distributions of individual characteristics. The first (second) row gives the mean (standard deviation) of the within-MSA marginal distribution over the 238 MSA's. As seen, MSA's are more diverse in their educational and racial compositions than in their age or gender composition. For example, the coefficient of variation (CV) across MSA's in the fraction of people without a high school degree is 0.43, but only 0.23 for the fraction of residents under the age of 25.

Table 2 reports statistics from the data on the equilibrium outcomes of the model. Specifically, it shows the cross-MSA mean, standard deviation and CV of the number of reported crimes per 1,000 population, the arrest rate (percent), the number of police per 1,000 population, the employment rate (percent) and within-MSA mean earnings. As seen, the mean of the crime rate across the MSA's is 56.5 with a standard deviation of 17.0. As reflected by the CV, the variation in the crime rate is similar to that of the arrest rate (0.36) and of the number of police (0.40).

TABLE 2
Summary statistics: outcomes

	Crime (per 1,000)	Arrest (%)	Police (per 1,000)	Employment (%)	Mean earnings (\$)
Mean	56.5	19.0	4.78	73.3	38,142
Std Dev.	17.0	6.85	1.91	6.87	7,340
CV	0.30	0.36	0.40	0.09	0.19

Number of Obs: 238 MSA's in the U.S.

The labour market outcomes, on the other hand, exhibit less variation across MSA's; the CV in the employment rate is 0.09 and that in earnings 0.19.

4. EMPIRICAL IMPLEMENTATION AND ESTIMATION

4.1. Additional empirical specifications

4.1.1. Distribution of individual types. To implement the model, we need to specify the relationship between individual observable demographics and unobservable type (l_n, η_n, κ_n) .²⁸ Given that MSA's differ in their population distribution of observables, type distributions will also differ across MSA's. Conditional on x , the l , η and κ are assumed to be independent. For an individual i with demographics x_i , regardless of the MSA, the probability that i 's human capital is of level l_m is given by

$$p_{l_m}(x_i) = \begin{cases} \Phi\left(\frac{\ln(l_m) - x_i' \lambda^l}{\sigma_l}\right) - \Phi\left(\frac{\ln(l_{m-1}) - x_i' \lambda^l}{\sigma_l}\right) & \text{for } 1 < m < N_l, \\ \Phi\left(\frac{\ln(l_m) - x_i' \lambda^l}{\sigma_l}\right) & \text{for } m = 1, \\ 1 - \Phi\left(\frac{\ln(l_{m-1}) - x_i' \lambda^l}{\sigma_l}\right) & \text{for } m = N_l, \end{cases} \quad (10)$$

The mass points l 's are assumed to be quantiles from $\ln N(\bar{x}' \lambda^l, \sigma_l^2)$, where \bar{x} is the national (population) mean of x .²⁹

The probability that i 's preference for engaging in crime is of level η_m is given by

$$p_{\eta_m}(x_i) = \begin{cases} \Phi\left(\frac{\eta_m - x_i' \lambda^\eta}{\sigma_\eta}\right) - \Phi\left(\frac{\eta_{m-1} - x_i' \lambda^\eta}{\sigma_\eta}\right) & \text{for } 1 < m < N_\eta, \\ \Phi\left(\frac{\eta_m - x_i' \lambda^\eta}{\sigma_\eta}\right) & \text{for } m = 1, \\ 1 - \Phi\left(\frac{\eta_{m-1} - x_i' \lambda^\eta}{\sigma_\eta}\right) & \text{for } m = N_\eta. \end{cases} \quad (11)$$

It can be shown that there exists an η^* such that people with $\eta \leq \eta^*$ will never commit a crime independent of equilibrium outcomes.³⁰ We, therefore, set the lowest η , $\eta_1 = \eta^*$. We also assume that the largest η (η_{N_η}) is such that individuals with η_{N_η} will always commit crimes.³¹ The other mass points η 's are assumed to be quantiles from the distribution $N(\bar{x}' \lambda^\eta, \sigma_\eta^2)$ that are above η^* .

28. We set $N_l = 20$, $N_\eta = 10$ and $N_\kappa = 10$ such that there are 2,000 types of individuals.

29. Heterogeneity in wages across cities arises from variation in x 's and from heterogeneity in human capital rental prices. Recall that we allow productivity (τ) to differ across cities. Given that we use only cross-sectional variation in wages, it would be difficult to identify city-specific unobservables in human capital levels and rental prices.

30. As is shown in Appendix A3, $\eta^* = \ln(l_1) + \kappa_1 + \ln(1 - \alpha) - \ln(\alpha l_{N_l} + l_1)$.

31. A zero-crime equilibrium is thus never possible.

The probability that i 's value of home option is of level κ_m , is given by

$$p_{\kappa_m}(x_i, j) = \begin{cases} \Phi\left(\frac{\ln(\kappa_m) - x_i' \lambda^\kappa}{\sigma_\kappa}\right) - \Phi\left(\frac{\ln(\kappa_{m-1}) - x_i' \lambda^\kappa}{\sigma_\kappa}\right) & \text{for } 1 < m < N_\kappa, \\ \Phi\left(\frac{\ln(\kappa_m) - x_i' \lambda^\kappa}{\sigma_\kappa}\right) & \text{for } m = 1, \\ 1 - \Phi\left(\frac{\ln(\kappa_{m-1}) - x_i' \lambda^\kappa}{\sigma_\kappa}\right) & \text{for } m = N_\kappa. \end{cases} \quad (12)$$

where the mass points κ 's are quantiles from the distribution $\ln N(\bar{x}' \lambda^\kappa, \sigma_\kappa^2)$.³²

4.1.2. Probability distribution of market equilibria. Given police size (s_j) , $q_j^n(s_j) = 0$ if $\Omega_j^n(s_j)$ is not supportable as a market equilibrium. As shown, all the supportable $\Omega_j^n(s_j)$'s can be ranked by their crime rates. Based on this fact, we assume that the probability a particular market equilibrium is realized depends on the ranking of its crime rate in the set of equilibria. It is not clear *a priori* whether a low crime or a high crime equilibrium is more likely, and we use the following flexible yet parsimonious structure that allows for various possible scenarios with only two parameters.³³

Let $N_j^*(s_j)$ be the number of market equilibria in city j under s_j , and $\{\tilde{\Omega}_j^{n'}(s_j)\}_{n'=1}^{N_j^*(s_j)}$ the set of these equilibria ranked from low crime rate to high crime rate. The probability of the n^{th} element in $\{\tilde{\Omega}_j^{n'}(s_j)\}_{n'=1}^{N_j^*(s_j)}$, i.e., the probability that the n^{th} -ranked equilibrium $\tilde{\Omega}_j^n(s_j)$ is realized is given by

$$q_j^n(s_j) = \frac{\exp\left[1 + \zeta_1(n_{j1}^* - n)I(n < n_{j1}^*) + \zeta_2(n - n_{j2}^*)I(n > n_{j2}^*)\right]}{\sum_{n'=1}^{N_j^*(s_j)} \exp\left[1 + \zeta_1(n_{j1}^* - n')I(n' < n_{j1}^*) + \zeta_2(n' - n_{j2}^*)I(n' > n_{j2}^*)\right]}, \quad (13)$$

where $I(\cdot)$ is the indicator function. If $N_j^*(s_j)$ is odd, $n_{j1}^*(s_j) = n_{j2}^*(s_j) = \frac{N_j^*(s_j) + 1}{2}$, that is, the median of $\{1, \dots, N_j^*(s_j)\}$. If $N_j^*(s_j)$ is even, $n_{j1}^*(s_j) = \frac{N_j^*(s_j)}{2} + 1$, the first number to the right of the median, and $n_{j2}^*(s_j) = \frac{N_j^*(s_j)}{2} - 1$, the first number to the left of the median. The two parameters, ζ_1 and ζ_2 , capture the relationship between the probability of an equilibrium and its ranking. For example, if $\zeta_1 = \zeta_2 = 0$, all equilibria are equally likely; if $\zeta_1 > 0$, $\zeta_2 > 0$, the distribution of equilibria is U-shaped; if $\zeta_1 < 0$, $\zeta_2 < 0$, the distribution of equilibria is inverse-U-shaped; if $\zeta_1 > 0$, $\zeta_2 < 0$, lower crime-rate equilibria are more likely, and if $\zeta_1 < 0$, $\zeta_2 > 0$, higher crime-rate equilibria are more likely.³⁴

4.2. Estimation

We estimate the model using simulated generalized method of moments. For each parameter configuration, we solve for the equilibria of the model, compute the model-predicted moments

32. We allow for an i.i.d. additive MSA-specific random effect, $\bar{\kappa}_j$, which is assumed to be drawn from the distribution $N(0, \sigma_\kappa^2)$.

33. There are a number of papers that have also estimated equilibrium selection rules jointly with the other model parameters. Examples of this approach include [Akerberg and Gowrisankaran \(2006\)](#), [Bajari et al. \(2010\)](#), [Bjorn and Vuong \(1984\)](#), [Card and Giuliano \(2013\)](#), and [Jia \(2008\)](#).

34. When one and only one ζ is zero, then all equilibria below or above the median are equally likely and the probabilities for others are monotonically increasing or decreasing with their rank depending on the sign of the other ζ .

for each equilibrium and integrate over them. The parameter estimates minimize the distance between the model-predicted moments ($M(\Theta)$) and the data moments (M^d):

$$\hat{\Theta} = \arg \min_{\Theta} \left\{ \left(M(\Theta) - M^d \right)' W \left(M(\Theta) - M^d \right) \right\},$$

where Θ is the vector of structural parameters, and W is a positive-definite weighting matrix.³⁵ Θ includes parameters governing the distributions over individuals in their human capital and tastes for crime and for home, the distributions of city-level unobservables, the arrest technology, the production technology, the return to crime, the consumption level if arrested, government preferences, and the probability distribution of market equilibria.

4.2.1. Identification. We provide the following intuition for identification of the model. First, the different outcomes across MSA's with the same observable characteristics arise from different unobservable characteristics and/or different equilibrium realizations. On the one hand, if there are no differences in unobservables, then all of these MSA's should have the same level of the police force, because the decision depends on the (common) expectation of outcomes over all possible equilibria. The other outcomes could be different, and "discretely" so, arising from different equilibria. For example, with the same level of the police force, crime rates across these MSA's will be distributed as "discretely" different mass points, because there are discrete types of citizens and hence discrete levels of potential crime rates.³⁶ On the other hand, if a unique equilibrium is always selected, then differences in all outcomes arise from MSA unobservables. Given the assumption that these unobservables are continuous random variables drawn from a smooth distribution, the observed differences in outcomes should be much smoother than the previous case. For example, given that a government now knows exactly which equilibrium will be realized, it is hard to rationalize that the same level of the police force is optimal across MSA's with "discretely" different crime rates. Between these two extremes, the parameters governing the distribution of the multiple equilibria and those governing the distribution of unobservables adjust so as to match the data.

Secondly, the reduced form of the model is a set of equations relating the equilibrium outcomes of the model in a given MSA to its distribution of population observables. In particular, each of the six equilibrium outcomes (the crime rate, the arrest rate, police expenditures, the employment rate, the mean wage, and the standard deviation of the wage) is a nonlinear function of the proportion of people in the MSA of each of the 192 different observable types. In the linear case, there would be 192×6 reduced form parameters which are functions of fifty-six structural parameters.³⁷ There are thus many within- and cross-equation restrictions. The moments we choose relate these equilibrium outcomes to the population proportions. Although the necessary condition for identification is satisfied, to provide some evidence beyond that, we have conducted Monte Carlo exercises in which we first simulated data with known parameter values, treated as the "truth" and then, using moments from the simulated data, started the estimation of the model

35. In particular, W is a diagonal matrix, the $(k, k)^{th}$ component of which is the inverse of the variance of the k^{th} moment, estimated from the data.

36. This argument relies on the assumption, as made in the model, that the distribution of unobservables has no mass point.

37. Based on a regression of each of the six equilibrium outcomes on the MSA-specific distribution of the 192 different observable types, in each case we found that the characteristics were jointly statistically significant. The F -statistics for the six outcomes were 23.6, 43.1, 13.9, 166.3, 134.4, and 89.8, with associated p -values of less than 0.00.

from a wide range of initial guesses of parameter values. In all cases, we were able to recover parameter values that are close to the “truth”. In addition, as in [Adda et al. \(2017\)](#), we show in Appendix A5 that our objective function is sensitive to discrete changes in each model parameter, suggesting that the objective function is at a (local) minimum.

4.2.2. Estimation routine. The estimation routine involves an outer loop searching over the parameter space, and an inner loop determining the set of equilibria. The following describes the inner loop for each city.

Step 1: Given a set of parameters Θ , calculate the mass points of l and κ as the quantiles from $\ln N(\bar{x}'\lambda^l, \sigma_l^2)$ and $\ln N(\bar{x}'\lambda^\kappa, \sigma_\kappa^2)$, respectively. Calculate the first mass point $\eta_1 = \eta^*$ as defined in Appendix A5, and other mass points η as quantiles from $N(\bar{x}'\lambda^\eta, \sigma_\eta^2)$ that are above η^* . Calculate the criminal propensity T_n according to equation (6), and the number of T_n types N^* as in Appendix A3. Index T_n such that $T_n \leq T_{n+1}$.

Step 2: For each city j , draw the city-level unobservable characteristics $(v_j, \tau_j, \epsilon_j, \bar{\kappa}_j)$. For each vector of observable characteristics x , which is assumed to be discrete, calculate the probability vectors $p_l(x)$, $p_\eta(x)$ and $p_\kappa(x, j)$, according to equations (10) to (12). Given the distribution of observable citizen characteristics $G_j(x)$, calculate the measure of each (l, η, κ) -type in city j . Derive the measure (h_{jn}) of each T_n -type in city j according to Appendix A3.

Step 3: Pick an s_j from the grid for the size of police force and solve for all market equilibria.³⁸ For each $n \in \{1, \dots, N^*\}$, carry out the following calculation: (1) suppose the crime rate $\mu_j^n = \sum_{n' \leq n} h_{jn'}$,

which happens when all $T_{n'}$ are criminals for $n' \leq n$ and all types $T_{n''}$ with $n'' > n$ choose between work and home according to condition (5). (2) The aggregate human capital employed (L_j^n) is given by the total human capital among those who choose to work as in (7). (3) Derive the rest of the components of $\Omega_j^n(s_j)$ as follows: given μ_j^n and s_j , calculate the arrest rate according to (2); given the choice portfolios specified in (1), calculate A_j^n as in equation (7) and r_j^n as in equation (1). (4) Calculate the term in the middle of the equilibrium condition (8). $\Omega_j^n(s_j)$ is an equilibrium if only if inequality (8) is satisfied.

Step 4: Calculate the government cost under s_j by integrating over potential equilibria, as in (9). Repeat Steps 3-4 until the optimal size of police force s_j^* and the associated set of market equilibria $\{\tilde{\Omega}_j^n(s_j^*)\}$ are found.

Step 5: Do this R times for every city $j \in \{1, \dots, 238\}$. Each of the simulated replica serves as the counterpart of a city in our model.

38. The grid points for s_j are parameter-city-specific. They are bounded between $[0, \infty)$, and they always include the points $\{s_j^n\}_{n=1}^{N^*}$ such that

$$T_{n+1} = (1 - \mu_j^n) \left[\pi_j^n \ln(c) + (1 - \pi_j^n) A_j^n \right] + \mu_j^n \ln \left(\frac{b_j^n}{1 - \alpha} \right) - \ln(r_j^n),$$

$$s.t. \pi_j^n = \Pi(s_j^n, \mu_j^n, \gamma_j, \epsilon_j).$$

That is, s_j^n is the level of the police force that makes type T_{n+1} at the margin between crime and decency (who chooses decency by assumption), and thus supports the equilibrium with μ_j^n exactly. Note that s_j^n may also support other equilibria, *i.e.* satisfying (8), to be determined in Step 3.

Step 6: Calculate the model predicted moments as

$$\frac{1}{238R} \sum_j \sum_{n=1}^{N^*} q_j^n(s_j^*) M_j^n(s_j^*; \Theta),$$

where $M_j^n(s_j^*; \Theta)$ is the vector of model predicted statistics in city j if $\Omega_j^n(s_j^*)$ is a market equilibrium.

4.2.3. Target moments. We target 171 moments.³⁹ These moments include, among others,

1. A set of unconditional moments with each MSA as an observation:
 - i) First moments across MSA's of the number of police, crime rate, arrest rate, employment rate, and the within-MSA average and standard deviation of earnings;
 - ii) Cross moments of the variables in 1), except that between mean earnings and the standard deviation of earnings;
 - iii) Second moments of the first 5 outcome variables.
 - iv) The fractions of cities with crime rates below the 10th, 20th, ..., 90th percentiles of crime rates in the data.
2. First moments of MSA outcomes by MSA characteristics.⁴⁰

The average size of the police force, the average crime rate and the average arrest rate conditional on the within-MSA marginal distributions of age, gender, education, and race, all treated as discrete variables. For example, we target the average outcomes among MSA's where the fraction of college graduates is ranked below the 50th percentile among all MSA's.
3. First moments of individual outcomes by individual characteristics.
 - i) The employment rate and average earnings among individuals by age, by education, by gender, by race, by the number of kids, and by the number of kids among females.
 - ii) It is well documented that the crime rate is significantly higher among youths. However, our data do not contain age-specific crime rates. To extract as much information as possible about the criminal versus non-criminal choices, we also use data from the CPS on school enrolment status. Specifically, we target, within this age group (less than age 25), their employment rate and school attendance rate by current education attainment.⁴¹

39. Although there are conceptually 192*6 potential moments, as discussed above, combining cells with small numbers of observations yielded the 171 moments we used in the estimation.

40. To keep the number of observations constant across moments that use MSA's as units of observations, we target joint moments of characteristics and outcome instead of conditional moments. For example, letting $I(\cdot)$ be the indicator function, we target $E(y_{jm}, I(G_j(x) \leq z))$ instead of $E(y_{jm} | G_j(x) \leq z)$, where y_{jm} is the m^{th} outcome, $G_j(x)$ is the distribution of within-MSA demographics, and z is some quantile. Similarly, in the next (the third) set of moments, which use individuals as units of observations, we target joint moments of individual outcome and individual characteristics, instead of conditional moments to keep the number of observations constant across this set of moments.

41. To incorporate school enrolment, we assume that, for people in the youngest age group, the home option includes attending school. With a probability that is specific to one's current education level, a youth who chooses the home option

TABLE 3
Selected parameter estimates

Production θ	0.84	(0.01)
Return to crime α	0.10	(0.003)
Consumption if arrested \underline{c}	0.09	(0.003)
Dispersion of MSA-level unobservables		
Productivity σ_τ	0.12	(0.09)
Arrest efficiency σ_ϵ	5.89	(0.03)
Value of leisure $\sigma_{\bar{c}}$	0.13	(0.07)
Marginal cost of police σ_v	0.25	(0.10)
Probability distribution of market equilibria		
Below median slope ζ_1	-0.20	(0.18)
Above median slope ζ_2	-0.60	(0.24)

4.3. Parameter estimates

Table 3 shows selected parameter estimates. Standard errors (in parentheses) are calculated via bootstrap.⁴² As shown in the upper panel of Table 3, the elasticity of output with respect to aggregate human capital (θ) in the legal sector is 0.84. The value of α , the return to crime, implies that a criminal steals 10% of the victim's income, or approximately \$3,250.⁴³ However, if apprehended, consumption is only \$900.⁴⁴

The middle panel of Table 3 shows the standard deviations of each of the four location-specific unobservables. To understand their magnitudes, we explore four counterfactual scenarios, where we increase, one at a time, legal sector productivity, the arrest efficiency, the value of leisure, and the marginal cost of police for every location by one standard deviation. We then calculate the percentage changes in outcomes at the national level as compared to the baseline. Without changing the number of police, the national crime rate is reduced by 6.3%, 17.7%, and 8.3% when labour productivity, police efficiency and the value of leisure are individually increased by one standard deviation above their current levels in all MSA's (see Appendix Table A1). Allowing for the government to optimally adjust the size of their police force leads to reductions in the number of police and thus smaller decreases in the crime rate of 2.8%, 11.9% and 3.6%. Finally, if the marginal cost of police is reduced by one standard deviation, the national police force increases by 14.8% leading to a 7.8% reduction in crime.⁴⁵

The lower panel of Table 3 shows the slope parameters that govern the distribution of market equilibria. Both parameters are estimated to be negative, although ζ_1 is not precisely estimated. In a Wald test, the p -value for the joint test that both of ζ 's are zero is 0.023. The point estimates suggest that for a given city's size of police force, the distribution of market equilibria is inverse-U-shaped. That is, the equilibrium whose crime rate is ranked in the middle is most likely to be realized. Moreover, as $|\zeta_1| < |\zeta_2|$, the distribution is asymmetric: given the same ranking distance from the

is enrolled in school. These probabilities are treated as parameters to be estimated jointly with other parameters in the model. For people of older ages, we restrict the home option to not include attending school, given that school enrolment rates are low for them.

42. There are 160 bootstrap iterations. Each iteration takes about 400 minutes to run using 48 CPU's and Message Passing Interface (MPI). In each iteration, we sample 238 MSA's from the original sample of 238 MSA's with replacement. All individuals in the matched CPS data are included who reside in MSA's in the bootstrap sample.

43. If they could choose only between work and home, *i.e.* criminal activity were not an option, only 31% of the current criminals would choose work over home, and their average income would be \$2,550. Obviously, criminals come from the lower tail of the income distribution.

44. All parameters in monetary units are in units of \$10,000.

45. These effects, as well as those that follow, are obtained by integrating over market equilibria.

TABLE 4
Means

	Crime	Arrest	Police	Employment	Mean wage	Std dev wage
	(%)	(%)	(%)	(%)	(\$)	(\$)
Data	5.65	19.03	0.48	73.33	38,142	40,067
Model	5.63	19.01	0.47	73.82	39,551	39,598

TABLE 5
Standard deviations

	Crime	Arrest	Police	Employment	Mean wage
	(%)	(%)	(%)	(%)	(\$)
Data	1.70	6.85	0.19	6.69	7,340
Model	1.53	7.79	0.12	4.34	5,996

median, the high crime rate equilibrium is less likely than the low crime rate equilibrium.⁴⁶ Based on our estimates, the number (N^*) of potential equilibria is 233, not all of which are supportable within an MSA. The modal number of equilibria across our simulated MSA's is 2, which happens in about 80% of the cases. In fewer than 0.3% of the cases, there is a unique equilibrium, which has been implicitly assumed in many studies. The average within-MSA dispersion of crime rates across equilibria is 1%, or more than half of the dispersion of crime rates across MSA's in the data (1.7%). In the counterfactual experiment section, we will illustrate the policy relevance of assuming a single equilibrium.

Appendix Tables A2–A4 show the parameters governing the distributions of values of the home option, levels of human capital, and tastes for crime conditional on x . Education is positively correlated with one's human capital and value of leisure. With the presence of young children, the value of leisure decreases (increases) for males (females), which is consistent with the fact that the employment rate for males (females) with children is higher (lower) when there are children in the household. The preference for engaging in crime is lower for those who are more educated, older and/or have young children.⁴⁷

4.4. *Model fit*

Table 4 shows the predicted means of the equilibrium outcomes across MSA's. As seen, they closely match the data. The fit to the standard deviations of outcomes across MSA's, as shown in Table 5, is generally close but not as good as the fit to the means.⁴⁸

Of particular interest are the correlations among the equilibrium outcomes. Recall that the observed outcomes in an MSA come from a particular equilibrium drawn from its set of equilibria. The collection of these data outcomes across the 238 MSA's leads to the observed correlations.

46. For example, if there are three market equilibria ranked by their crime rates from low to high, then the probability distribution will be {0.35, 0.42, 0.23}. In the case of four market equilibria, the distribution will be {0.29, 0.35, 0.23, 0.13}.

47. Appendix Table A5 shows the probabilities of attending school among youth who choose the home option. Table A6 shows estimates for government preferences and other parameters in the arrest technology.

48. In the following, unless otherwise stated, we measure the crime rate in percentage of population, that is, crimes per 100 population, instead of per 1,000 population. This measure is more consistent with our model, where the crime rate is defined as the percentage of citizens who choose to be criminals.

TABLE 6
Correlations between outcomes

	Data		Model	
	Mean	Std.Dev.	Mean	Std.Dev.
(Police, Crime)	0.35	0.16	0.23	0.06
(Police, Arrest)	-0.08	0.08	-0.04	0.06
(Police, Employment)	-0.18	0.09	-0.12	0.06
(Police, Average Wage)	0.18	0.13	-0.06	0.10
(Employment, Crime)	-0.30	0.14	-0.38	0.06
(Employment, Arrest)	0.10	0.09	0.20	0.08
(Employment, Average Wage)	0.06	0.07	0.18	0.09
(Crime, Arrest)	-0.20	0.10	-0.39	0.10
(Crime, Average Wage)	-0.16	0.09	-0.25	0.10
(Arrest, Average Wage)	-0.11	0.08	0.01	0.07

TABLE 7
Correlation between crime rates and MSA-level demographics

	Data	Model
P(Race=Black/Hispanics)	0.41	0.40
P(Education<HS)	0.26	0.38
P(Education>=College)	-0.21	-0.31

Using our model, we can simulate a large number of cross-sectional observations. One cross-section consists of 238 MSA's, each with a particular equilibrium drawn from its set of equilibria.⁴⁹ For each cross section, we can calculate one set of correlations. Across all simulated cross sections, we can obtain the model distribution of correlations. The results are shown in Table 6. The first column shows the correlations in the data. The second column shows the standard deviations of these correlations obtained via bootstrapping. The third (fourth) column shows the mean (standard deviation) of the correlations from the simulation. In all but two cases (out of ten) the model correctly predicts the sign of these correlations. For example, the correlation between the number of police and the crime rate is positive both in the data and for the model prediction. The positive correlation arises from two offsetting effects. On the one hand, having additional police leads to a reduction in crime through an increase in the apprehension rate. On the other hand, governments in MSA's that can be expected to be more prone to crime because of the distribution of citizen characteristics and/or MSA-level characteristics will optimally choose a larger police force. When the model and the data diverge in sign, between police size and the wage and between the arrest rate and the wage, the correlations in both the data and the model tend to be weak and variable. In addition, the differences between the data and model predictions are generally small in magnitude.

The model fits well the correlations between the crime rate and the within-MSA distribution of demographics, as shown in Table 7. The correlation between the crime rate and the fraction of blacks/Hispanics is 0.41 in the data and 0.40 in the model and that between the crime rate and the fraction of citizens with less than high school (at least college) education is 0.26 (-0.21) in the data and 0.38 (-0.31) in the model.⁵⁰

To further assess our estimated model, we use data from a different year than our estimation sample to perform an out-of-sample validation. In particular, we use the year 2003, for which

49. This exercise still requires one solve for all possible market equilibria.

50. Tables A13–A15 show more model fits.

we are able to match 192 MSA's between the CPS and the UCR.⁵¹ Appendix Tables A16–A18, which are the counterparts for Tables 4–6, show that the model is able to replicate the patterns in 2003. For example, compared to the 2008 sample, on average, the 2003 sample has a higher crime rate (6.32% versus 5.65%) and a lower arrest rate (19.0% versus 17.0%). The model is able to capture these differences and predicts the crime rates for the two different samples as 5.98% versus 5.63% and the arrest rates as 19.0% versus 18.2%.

4.4.1. Comparing equilibria. Our validation exercise assumed that the equilibrium selection rule, *i.e.* the equilibrium probability distribution function $q_j^n(\cdot)$, is the same over time and that equilibria are selected independently over time. Having multiple years of data raises an interesting question as to whether selected equilibria are correlated over time. To answer this question requires that we be able to compare equilibria from different equilibrium sets to determine their “sameness”. Such a comparison is complicated because equilibrium sets may differ in the number of equilibria and in their equilibrium outcomes. A natural way to compare equilibria is by their rank in terms of the crime rate within its equilibrium set, accounting for the different number of equilibria in the set. Specifically, we adopt the following definition of “sameness”. Within a set, the rank of an equilibrium is relative to three positions, the lowest-crime-rate equilibrium (e_l), the highest-crime-rate equilibrium (e_h), and the median in that set. The rank of an equilibrium below (above) the median is determined by its position-wise distances from e_l (e_h) and the median. Two equilibria, either over time in the same MSA or at a point in time across MSA's, are deemed to be the same if they have the same rank within their equilibrium set.

As noted, the external validation described above assumed that the equilibrium selection rule in 2003 and 2008 were the same and that the equilibrium selected in the 2 years were drawn independently. However, because our estimated equilibrium selection rule implies that equilibria are not uniformly distributed, independent draws may still lead to a higher probability that the same equilibrium was selected in the 2 years. Indeed, we find that the same equilibrium was selected with 77% probability.⁵²

5. EXPLAINING CRIME RATE DIFFERENCES ACROSS MSA'S

In this section, we examine the role of MSA observable characteristics and that of multiple equilibria in explaining the difference in crime rates across MSA's. Because we explicitly model the distribution of multiple equilibria, we can use our model to recover the combinations of unobservable characteristics and equilibrium selection that could have generated the data for each MSA. To isolate the effect of those factors, we pair 110 relatively large MSA's each with a single comparison MSA, Philadelphia.⁵³ We decompose the pair-wise difference in outcomes in three steps.⁵⁴

51. The MSA code changed between 2003 and 2008, leading to fewer matches.

52. We also estimated the model with data from both 2003 and 2008, imposing the restriction that the same equilibrium had been selected in both years within an MSA. This exercise, suggested by the editor and a referee, is intended to see whether the results are robust to a further restriction on the equilibrium selection rule. Note, however, that such an assumption would be consistent with our model, in particular, with the objective function of the government, only if the government either did not recognize that the same equilibria is selected or if the government is myopic. Allowing for state-dependence in the equilibrium selection rule would require a fully dynamic model in which the government accounts for the dependence, which is beyond the scope of this article. Appendix A6 provides results of the analysis.

53. We use 110 pairs to guarantee enough observations while taking into account the computational burden. The choice of Philadelphia as the baseline MSA is illustrative (though not random).

54. See Appendix A2 for more details.

TABLE 8
Crime rate comparison (Philadelphia versus Others)

%	Base ^a (μ_j^{BASE})	Counterfact ^b (μ_j^{PHL})	Philadelphia ^a (μ_{PHL}^{PHL})
Phoenix-Mesa-Scottsdale	6.9	5.9	5.3
Atlanta-Sandy Springs-Marietta	6.1	5.2	5.1
Houston-Sugar	6.5	6.4	5.3

^aFor each pair of MSA's, compute the equilibrium that best matches the data for each MSA under a common vector of acceptable unobservables, then average over a large number of such vectors.

^bUnder the same unobservables as in *a*, Row MSA selects the same equilibrium from its set as the one in Philadelphia's baseline equilibrium.

Step 1: To control for unobservables, we find a vector of unobservable characteristics common to both MSA's in a pair that can generate equilibrium outcomes (the crime rate, the level of police, the arrest rate, the employment rate, and the average wage) matching those observed in each MSA of the pair. We do this for each pair. Given that each pair has the same vector of unobservables, the resulting equilibrium crime rates (μ_j^{BASE}), one for each MSA in the pair, differ between a pair only because of differences in their observable characteristics and in their equilibrium selection.

Step 2: To examine the role of equilibrium selection, we compare an MSA's baseline crime rate (μ_j^{BASE}) with the counterfactual crime rate that would have resulted had the "same" equilibrium, as defined in last section, been selected as the one in Philadelphia (μ_j^{PHL}).

Step 3: To examine the role of observables in explaining the cross-MSA differences in crime rates, we consider the relationship between observables and the crime rate gap between an MSA and Philadelphia when both use the Philadelphia equilibrium selection rule (μ_j^{PHL}).

5.1. Results

The data-matching equilibrium crime rates (μ_j^{BASE}) are shown in the first column of Table 8 for each of the row MSA's and in the last column for Philadelphia. We show three example MSA's for illustration. Given that we use the same unobservables for a pair, the differential crime rates between an MSA and Philadelphia arise from the difference in their observable characteristics and the difference in their equilibrium selection. As seen, the difference in crime rates due to those factors are 1.6, 1.0, and 1.2 crimes per 100 population in the three comparisons.

If the same equilibrium as the one in Philadelphia was selected, the crime rate (μ_j^{PHL}) would be lower than the baseline for 76% of the 110 MSA's.⁵⁵ For the three example MSA's, the counterfactual μ_j^{PHL} is reported in the second column of Table 8. The crime rates in Phoenix and Atlanta would decrease significantly if they used the Philadelphia equilibrium selection rule. In particular, the crime rate in Atlanta would fall almost to the same level as that in Philadelphia. In contrast, the differential crime rate between Houston and Philadelphia is hardly driven by the different equilibria that are selected.⁵⁶

Finally, we examine the role of observables in explaining the crime rate gap between MSA's and Philadelphia when the MSA's have the same equilibrium as Philadelphia, *i.e.* Column 2 minus Column 3 of Table 8 ($\mu_j^{PHL} - \mu_j^{PHL}$). We run a linear regression using the 110 pairs

55. The mean (standard deviation) of ($\mu_j^{PHL} - \mu_j^{base}$) across the 110 MSA's is 1.73% (2.98%).

56. Although using cities other than Philadelphia as the baseline will likely change the quantitative results, Philadelphia is instructive because of the variety of patterns that are observed.

TABLE 9
Crime rate and observables

obs: 110 MSAs	$\mu_j^{PHL} - \mu_{PHL}^{PHL}$		$\mu_j^{data} - \mu_{PHL}^{data}$	
	Estimate*	Std error	Estimate	Std error
P(HS dropout)	0.04	(0.01)	0.03	(0.03)
P(HS grads)	0.04	(0.01)	0.04	(0.02)
P(Some college)	0.01	(0.01)	0.04	(0.02)
P(Age ≤ 25)	0.11	(0.02)	0.01	(0.04)
P(26 < Age ≤ 40)	0.05	(0.01)	-0.01	(0.04)
P(40 < Age ≤ 50)	-0.02	(0.02)	-0.04	(0.04)
P(Male)	0.04	(0.02)	-0.01	(0.04)
P(Black or Hispanic)	0.02	(0.004)	0.04	(0.01)
P(no child)	0.11	(0.02)	0.09	(0.05)
P(one child)	0.03	(0.03)	0.04	(0.07)
Constant	-0.17	(0.03)	-0.10	(0.07)
R ²	0.82		0.37	

*MSA unobservables included in the regression $\mu_j^{PHL} - \mu_{PHL}^{PHL}$.

of $(\mu_j^{PHL} - \mu_{PHL}^{PHL})$ on MSA j 's observable and unobservable characteristics.⁵⁷ We contrast that regression with one based on the data, that is, not controlling for unobservables and equilibrium selection. The results are shown in Table 9, where each row variable is the fraction of population of that category in an MSA.⁵⁸ Controlling for unobservables and the selection of equilibria, an MSA is more prone to a high crime rate if it has higher fractions of lower-educated people (high school and below), younger people (below age 40), males, blacks, or Hispanics, and people without children. The high R^2 (0.82) reflects the conditioning on unobservables and equilibrium selection. The lack of a perfect fit is because the same equilibrium in rank may correspond to different outcomes and because the relationships may not be linear. Residual variation in those regressions include variation in unobservables and in equilibrium selection. The resulting R^2 is thus considerably smaller (0.37). Many of the variables in the regression without the conditioning have large standard errors, which makes comparisons between the regressions problematic. With that caveat, concentrating on point estimates, in ten of the twelve cases, the signs are the same. In terms of magnitudes, the greatest differences are in the effect of age on crime, which declines less without conditioning, and male gender.⁵⁹ Given orthogonality between the observable variables and the unobservables, differences in these coefficient values arise from differences in equilibrium selection. The general point is that regression estimates based on the data would not represent the effect of the observables on crime.

57. Given the nonlinear relationship between crime rates and an MSA's observable and unobservable characteristics, this linear regression is only meant as an approximation to the relationship.

58. These characteristics are considered as important crime-related factors in the literature. For example, Levitt and Lochner (2000) find factors such as gender, family environment, and cognitive ability are good predictors of criminal involvement. Lochner and Moretti (2004) find that school attainment significantly reduces participation in criminal activity. Other studies have explored the extensions of the mandatory schooling age or the cut-off birth date for enrolment to study the effect of education on crime (Machin *et al.*, 2011; Clay *et al.*, 2012; Anderson, 2014; Hjalmarsson *et al.*, 2015; and Cook and Kang, 2016).

59. Some of these characteristics, such as age, education, and numbers of children, change over the life cycle. An implicit assumption of our framework is that MSA's are composed of overlapping generations, that none of these life cycle changes are choices or affected by the past choices in the model, and that MSA's are in a stationary equilibrium.

TABLE 10
National outcomes

	Uniform	Subsidy 1	Subsidy 2	Subsidy 3
% MSA with increase in police	100	92.3	65.8	77.8
% Change in total police	20	41.9	20	25.6
% Change in crime rate	-8.23	-21.2	-10.6	-13.4
Elasticity crime to police	-0.41	-0.51	-0.53	-0.53
Federal-funded additional police per 1,000 people	0.94	0.94	0.25	0.40
Total additional police per 1,000 people	0.94	1.97	0.94	1.20
$\frac{\% \text{Reduction in crime rate}}{\text{Federal-funded additional police per 1,000 people}}$	8.80	22.7	41.9	33.4
$\frac{\% \text{Reduction in crime rate}}{\text{Total additional police per 1,000 people}}$	8.80	10.8	11.3	11.2

6. COUNTERFACTUAL POLICY EXPERIMENTS

The counterfactual experiments are motivated by federal programmes that have provided subsidies to local governments for increasing the number of police. For example, the COPS programme, initiated by the Clinton administration in 1994 and still on going today, planned to add 100,000 (20%) more police nationwide by fiscal year 2000, as part of the federal government's effort to reduce crime.⁶⁰

6.1. *Increasing the Police Force Nationwide*

The first counterfactual experiment considers scenarios where the planner (federal government) aims at an increase of 20% in the total number of police through a subsidy to local governments and where the use of the subsidy is monitored. First, we consider a scenario where the programme is "uniform" in the sense that the size of the police force is increased by 20% in each MSA.⁶¹ Secondly, we consider three scenarios where the planner subsidizes a certain percentage of the newly hired police. However, the number of new hires in each MSA is chosen by each local government, that is, the take-up of the subsidy is voluntary. The three scenarios differ in the subsidy rate.

Table 10 shows the outcomes of these four scenarios at the national level. The first row shows the percentage of MSA's that experience an increase in their police force. Rows 2–4 show the percent change nationwide in the total police force, the percent change in the crime rate and the implied elasticity of the crime rate with respect to the number of police. Row 5 shows the number of additional police per 1,000 people funded by the federal government. Row 6 shows the number of additional police per 1,000 people funded by both the federal and the local governments. The next two rows show the efficiency of each policy, as measured by the ratios of the reduction in crime to Row 5 ("federal efficiency ratio") and to Row 6 ("overall efficiency ratio").

Under a uniform policy (column 1), the police force increases by 20% and the crime rate falls by 8.23%, *i.e.* the elasticity of crime rate with respect to the number of police is -0.41. The planner funds an additional 0.94 police officers per 1,000 people and for each unit of federal resources, as measured by the number of additional police per 1,000 people, there is a 8.8%

60. Since 1994, COPS has provided \$11.3 billion in assistance to state and local law enforcement agencies to help in hiring additional police officers.

61. This increase in police is 0.5 of a standard deviation, well within sample variation.

reduction in crime, that is, both efficiency ratios are 8.8, since all new hires are funded by the federal government.⁶²

To illustrate the potential biases arising from ignoring the existence of multiple equilibria, we re-estimated the model under the restriction that when multiple equilibria are possible, the one with the highest (lowest) crime rate is always selected. Appendix Table A7 contrasts the predicted effects of the uniform policy under these two alternative models with the one shown in Column 1 of Table 10. Compared to the -0.41 elasticity predicted by the unrestricted (original) model, the high-crime-only model predicts a crime-to-police elasticity of -0.61 , and the low-crime-only model predicts an elasticity of -0.09 . Although the two equilibrium selection rules we have imposed may be extreme, the exercise is not meant to argue for or against a particular “single equilibrium” selection rule. Rather, it illustrates that the choice of an equilibrium selection rule can be consequential in a policy analysis of this kind.

The next set of scenarios requires that the subsidy rate lead to outcomes that satisfy three different constraints. Because the hiring decisions of local governments vary with the subsidy policy, we solve for the desired subsidy rate by simulation in each case. We simulate equilibrium outcomes for a given subsidy rate and repeat the process until we find the subsidy rate that satisfies our constraint. In the first scenario (column 2), we find the subsidy rate such that the number of federally funded additional police is the same as in column 1, that is, the employment of 0.94 additional police per 1,000 people. That rate turns out to be 47.7%. At that subsidy rate, over 92% of local governments accept the subsidy and increase their police force. The number of police increases by 41.9% nationwide, considerably greater than the target increase of 20%, and the national crime rate decreases by 21.2%. The federal efficiency ratio is 22.7, more than twice that of the uniform policy. The overall efficiency ratio is 10.8, about 23% higher than that of the uniform policy.

In the next scenario, we find the subsidy rate such that the nationwide increase in total police is the same as the planner intended, that is, 20%. That subsidy rate turns out to be 27%, the number of federal-funded police officers is 0.25 per 1,000 people and each unit of federal resources leads to an almost 42% decrease in crime. Notice that the total number of additional police is the same in this scenario as in the scenario with the uniform 20% nationwide increase in the police force. However, in this case there is a more significant reduction in crime (10.6% versus 8.2%). With monitoring, as is assumed in both cases, allowing local governments to choose their own police force, instead of a uniformly imposed police increase, leads to larger reduction in crime nationwide and an overall efficiency ratio at 11.3.

In the last scenario, we consider the effects of a COPS-like budget on crime reduction. We first calculate the intended cost of the COPS programme. With the 75% COPS subsidy rate and a goal of 100,000 new police, the cost is equivalent to 0.4 federal-funded police per 1,000 people in the year 2000. The subsidy rate that satisfies this budget constraint is 34%. About 78% of local governments take the offer, leading to a 26 percent increase in police nationwide, higher than the 20% increase set as a goal by COPS. The crime rate is reduced by 13%, the federal (overall) efficiency ratio in this case is over 33 (11). These ratios are what one could expect from a COPS-like programme if the federal government were able to monitor local governments perfectly so

62. Using the bootstrapped parameters, we have calculated the standard deviation of the policy impact. At our estimates, the changes in the employment rate, the crime rate, and the arrest rate are, respectively, 0.47%, -8.23% , and 28.02%. The standard deviations are, respectively, 0.07%, 1.34%, and 1.37%.

TABLE 11
Distribution of outcomes across MSA's

	Uniform		Subsidy 2	
	Std dev	CV	Std dev	CV
% Change in police	0	0	20.17	0.98
% Change in crime rate	3.86	0.46	8.79	0.87

that the grants are used exclusively to add new police on net. That is, this scenario provides an upper bound on the effects of such programmes.⁶³

The effects of subsidies can differ substantially across locations. For example, Table 11 shows the standard deviations and coefficients of variation of various outcomes across MSA's under the two scenarios represented in columns 1 and 3 of Table 10, where the total increases in police are both set at 20%. In the first column, although the percentage change in police is uniform across MSA's, the reduction in crime is dispersed with a CV in the crime rate of about 0.5. In column 2, the CVs of both the increase in police and the change in crime are not far from one.

6.1.1. Simulating the policy impact using model-generated IV. There is an extensive regression-based literature that attempts to estimate the impact of increasing police on crime. First generation papers were based on OLS regressions. However, an OLS regression of crime on police, say across MSA's, would confound the effect of increasing police on criminal behaviour with the response of police to increased crime, as in our equilibrium model. Indeed, as seen in Table 6, the cross-sectional correlation between crime and police in our data is positive. Recognizing this fact, second-generation papers adopted an instrumental variables approach.⁶⁴ There is wide variation in the estimates of the elasticity of crime to police depending on the instruments used and on the types of crimes considered.⁶⁵ However, as noted by Chalfin and McCrary (2017), few of the estimated elasticities are significant, reflecting a great deal of sampling variability and the use of relatively weak instruments.⁶⁶ Draca *et al.* (2011), using the shift in police deployment following the July 2005 terror attacks in London as an exogenous event, find the crime-to-police elasticity to be around -0.38 based on the overall crime rate.

Both the existence of multiple equilibria and the interpretation of the IV estimate as a local average treatment effect (LATE) make it difficult to compare these IV estimates to that generated by the counterfactual we report in Table 10. To explore that issue, we use our model to conduct an IV estimation of the effect of police on crime, where the IV is an exogenous reduction in the cost of police, in the spirit of the COPS programme. In particular, we compare crime rates under the baseline given by our parameter estimates to a hypothetical in which an exogenous reduction in the marginal cost of police, v as in Equation (9), increases the size of the police force. We implement this exercise in the following way. First, we simulate the baseline data by drawing a

63. Evans and Owens (2007) find that each officer funded by the COPS grant led to a 0.7 increase in police force, suggesting that the use of the COPS grant was imperfectly monitored. They also find a 10% decrease in property crime for each additional police officer per 1,000 people, that is, a federal efficiency ratio of 10.

64. See Worrall and Kovandzic (2010) for references to both first and second-generation papers.

65. For example, the estimate is -0.5 for property crime in Levitt (2002) using firefighter numbers as an IV, -0.26 for property crime and -1.34 for robbery in Evans and Owens (2007) using the size of COPS grants, -0.9 for property crime in Lin (2009) using the differential exposure to exchange rate shocks.

66. Chalfin and McCrary (2013) use measurement error corrected IV and obtain relatively precise estimates of elasticities at -0.17 for property crime and -0.56 for robbery.

TABLE 12
Wald estimates of police on crime: decrease in marginal costs of police

	Same-Ranked Equilibrium			Random Equilibrium
70% decrease	50% decrease	20% decrease	10% decrease	10% decrease
-5.86 (0.21)	-5.52 (0.39)	-6.66 (0.92)	-6.80 (1.96)	-3.63 (1.90)

Units: crime per 1,000 population, police per 1,000 population.

set of unobservables for each of the 238 MSAs and drawing an equilibrium for each MSA from the equilibrium set according to the estimated equilibrium selection rule. Secondly, using the same unobservables, we calculate the level of police and the crime rate resulting from a given reduction in ν under the same-rank equilibrium as the one in the baseline. We then calculate a Wald estimate of the effect of police on crime, given by the ratio of the difference in the mean crime rate over MSA's due to the reduction in ν to the difference in the mean level of police over MSA's due to that same reduction in ν . This process generates one Wald estimate based on a single realization for the set of 238 MSA's before and after the reduction in the cost of police. We do this 1,000 times and obtain a distribution of Wald estimates.⁶⁷

Table 12 reports the results. The Wald estimate is interpreted as the effect on the crime rate (crimes per 1,000 population) of a one unit increase in the number of police per 1,000 population, a 21% increase at the sample mean. We present four sets of the mean and standard deviation (in parentheses) of the Wald estimates using the same ranked equilibrium for the hypothetical (the first four columns of Table 12), each set calculated over the 1,000 samples. The IV in each of the four sets is a different size reduction in the marginal cost of the police. The mean of the Wald estimate ranges from -5.52 when there is a 50% reduction in ν to -6.80 when there is a 10% reduction ν .⁶⁸ The LATE estimate thus seems to be fairly stable with respect to the size of the reduction. However, the standard deviation of the Wald estimate is considerably more sensitive. As seen in the table, the standard deviation monotonically decreases with increases in ν ; it ranges from a high of 1.96 to a low of 0.21 as ν increases from 10% to 70%. This fall in the standard deviation indicates that as ν becomes large, the LATE estimate is closer to the average treatment effect (ATE). Of course, in actual applications, one would not be able to tell if LATE and ATE are close. The general point, as seen in the simulations, is that these IV estimates strictly apply only to a particular instrument (the level of ν), and, unless the sample is "large", to a particular sample (a particular realized set of unobservables).

To illustrate the role of multiple equilibria in IV estimation, we repeat the exercise under the 10% reduction in ν , where we randomly draw an equilibrium after the reduction instead of imposing that the post-reduction equilibrium be the same rank as the one in the baseline. The last column of Table 12 shows that the mean of the Wald estimate under the random selection rule is -3.63, as compared to -6.80 under the same-equilibrium selection rule. The magnitude of the LATE estimate is thus also significantly affected by the equilibrium selection.

67. The variance in the Wald estimates understates the true variance because we have fixed the parameters of the model at the original estimates. To obtain the true distribution, we would have to re-estimate the model parameters for each of the 1,000 draws and calculate each Wald estimate at the relevant parameter values for that sample.

68. The former amounts to a 9.8% decrease in the crime rate at the mean over all MSA's; and the latter amounts to a 12% decrease, implying that the estimated elasticity ranges from -0.47 to -0.57. Notice that although these elasticities are in the range of those reported in Table 10, they are not strictly comparable. Table 10 refers to the national crime rate, while Table 12 is at the MSA level.

6.2. Targeted allocation of the police force

Given that the effects of additional police on crime differ significantly across MSA's, determining the efficient allocation of extra resources across MSA's becomes a nontrivial task. To avoid excessive computational burden, we focus on pair-wise MSA police allocations, although it is conceptually feasible to consider larger sets of MSA's. Our model can be used to examine the effect of a wide range of police allocations across MSA's. To provide a policy context, we examine three different objective functions for the policymaker. As researchers, it is not our role to act as policymakers, but rather to determine the impact of alternative policies in achieving the policymaker's objective. We take these objective functions simply as illustrative.

To be concrete, we take as the policy a 20% overall increase in the police force which is to be allocated across two MSA's. The first objective function of the policymaker is to minimize the mean (weighted by population) of the expected (taken over the set of equilibria) crime rates over the two MSA's.⁶⁹ The second, presented in Appendix Tables A9–A10, is to minimize the mean (weighted by population) crime rate where the crime rate comes from the most-likely equilibrium (as opposed to taking the expectation over equilibria). The third, contained in the online Appendix, is to minimize the mean (weighted by population) crime rate, where the crime rate comes from the highest-crime-rate equilibrium.

We consider five different police allocation rules.

Rule 1 (Complete Information): the policymaker has perfect information on both the MSA-specific observable and unobservable characteristics and chooses the police allocation that satisfies the objective function.⁷⁰

Rule 2 (Based on x): the policymaker has information only on the observable characteristics of each MSA and chooses the police allocation that satisfies the objective function.

Rule 3 (Based on current crime rate): the policymaker does not satisfy an objective function but instead allocates police according to the ratio of the population-size-adjusted crime rates, such that the high crime rate MSA receives more extra police per capita. Concretely, letting pop_j and μ_j be the population size and crime rate in MSA $_j$, the ratio of the police allocation across the two MSA's is given by $\frac{p_1}{p_2} = \frac{\mu_1 pop_1}{\mu_2 pop_2}$.

Rule 4 (Based on current arrest rate): the policymaker does not satisfy an objective function but instead allocates police according to the size-adjusted inverse arrest rate ratio, such that the low arrest rate MSA gets more extra police per capita. Let π_j be the arrest rate in MSA $_j$, the ratio of police allocation is given by $\frac{p_1}{p_2} = \frac{\pi_2 pop_1}{\pi_1 pop_2}$.

69. Let $E(\mu_j|p, u)$ be its expected crime rate under allocation p and unobservable vector u . In the case of complete information, rule 1 below, where the planner knows the unobservable of MSA j , the goal is to

$$\min_p \left\{ \sum_{j=1}^2 \frac{pop_j}{pop_1 + pop_2} E(\mu_j|p, u_j) \right\}.$$

In the case of incomplete information, rule 2 below, an additional outer layer of expectation is taken over unobservable characteristics; and the goal is to

$$\min_p \left\{ \sum_{j=1}^2 \frac{pop_j}{pop_1 + pop_2} \int_u E(\mu_j|p, u) dF(u) \right\}.$$

70. Information on observable characteristics refers to the joint distribution of the whole vector of x 's, not just the marginal distribution of its components.

TABLE 13
Observable MSA characteristics

(%)	Population	Education		Age		Race Black/Hisp.	Gender Male
		Less than HS	BA or above	≤25	>50		
Philadelphia	2.38	11.45	31.43	18.91	27.16	26.01	47.66
Phoenix	2.01	19.61	25.34	19.69	26.09	32.89	50.12

TABLE 14
Current outcomes

	Philadelphia		Phoenix	
	Data	Model	Data	Model
Police (per 1,000)	4.66	4.66	5.01	4.80
Crime (%)	4.91	5.19	6.92	6.89
Arrest (%)	22.05	21.96	14.03	12.94
GDP per capita (\$)	33,297	32,681	28,807	29,073

Rule 5 (Based on current GDP): the policymaker does not satisfy an objective function but instead allocates police according to the size-adjusted inverse GDP per capita ratio, such that the low-GDP per capita MSA gets more extra police per capita. Let I_j be GDP per capita in MSA_{*j*}, the ratio of police allocation is given by $\frac{p_1}{p_2} = \frac{I_2 pop_1}{I_1 pop_2}$.

Although they may not be optimal, the last three allocation rules can be seen as representing a reasonable rule-of-thumb and appear to be realistic in practice as well. For example, [Evans and Owens \(2007\)](#) find that the COPS grants received by local governments are positively correlated with the local crime rate, the population size, the fraction of youths and the fraction of blacks, the latter two being components of x in our model.

As an example, we consider the allocation problem between Philadelphia and Phoenix-Mesa-Scottsdale. Table 13 shows the demographic characteristics of these two MSA's. The first column shows the population size of each MSA as a percentage of the total population size of the 238 MSA's in our data. Philadelphia is larger than Phoenix, representing 2.38% of the population versus 2.01%. Compared to Phoenix, educational attainment is higher and the fraction of black and/or Hispanics is lower in Philadelphia; the age and gender composition in these two MSA's are similar.

To implement Rule 1 (complete information), we first uncover the unobserved MSA-specific characteristics, based on our parameter estimates, by searching over the space of unobservable characteristics such that the distance between the model predicted and data outcomes is minimized MSA by MSA. In particular, for each MSA, we seek the combinations of MSA-specific arrest efficiency (ϵ), mean leisure value (\bar{k}), marginal cost of police (ν) and productivity (τ) to minimize the percentage model-data discrepancy over the crime rate, arrest rate, police force, employment rate, and average wage. Table 14 shows the outcomes in the data and those predicted by our baseline model with the sets of discrepancy-minimizing unobservables. Philadelphia has fewer police per-capita, a lower crime rate, a higher arrest rate and a larger per capita GDP than does Phoenix.

As shown in Table 13, Philadelphia has a less crime-prone population (*e.g.* fewer high school dropouts, fewer blacks) than Phoenix. However, their different outcomes are also due

TABLE 15
Unobservable MSA characteristics (std dev from the overall MSA sample mean)

	Productivity	Value of leisure	Arrest efficiency	Marginal cost of police
Philadelphia	0.54	1.46	0.15	-0.06
Phoenix	-0.15	-2.21	-0.49	-1.40

TABLE 16
Police allocation and crime reduction: Philadelphia versus Phoenix

%	Extra Police for Philly	Crime Reduction		
		Overall	Philly	Phoenix
Complete information	91.0	16.0	21.8	9.7
Based on x	98.0	10.1	21.8	0.0
Based on crime rates	47.2	5.6	0.1	10.7
Based on arrest rates	40.8	5.6	0.1	10.7
Based on GDP	51.4	7.0	3.0	10.7

to differences in unobservable population characteristics. Table 15 displays the unobservables of these two MSA's that we recover using our model, measured as standard deviations from the population mean over all MSA's. Productivity in Philadelphia is about half of a standard deviation above the population mean and the value of leisure is 1.5 standard deviations above the mean. Philadelphia also has slightly more efficient but a lower opportunity cost of police relative to the mean. In comparison, Phoenix has slightly lower-than-average productivity, a significantly lower-than-average value of leisure, arrest efficiency and marginal cost of the police. As mentioned in the parameter estimate section and shown in Table A1, higher productivity, higher arrest efficiency, a higher value of leisure, and a lower marginal cost of police each contributes to a lower crime rate conditional on observables. Except for its lower marginal cost of police, all of its characteristics make Phoenix more likely to have a high crime rate than Philadelphia.

We present the results of our allocation experiments for the first objective function in Table 16 for each allocation rule. The first column shows the share of extra police allocated to Philadelphia. The second column shows the associated percentage crime reduction in the overall crime rate, *i.e.* the policy goal. The next two columns show the reduction in each MSA. With complete information, 91% of the extra police is allocated to Philadelphia and the crime rate is reduced by 16% overall, by 22% in Philadelphia, and by 10% in Phoenix. Based on observable characteristics only, the planner allocates 98% of the extra police to Philadelphia, leading to a 10% reduction in overall crime, zero reduction in Phoenix and the same 22% reduction in Philadelphia as in the previous case. Based on the three separate outcome variables, the crime rate, the arrest rate, and GDP, about 40% to 50% of police would be allocated to Philadelphia. Reductions in the overall crime rate one-half of what it would be if the allocation is based on observables alone and one-third if based both on observable and unobservable MSA characteristics, mostly due to much lower reductions in Philadelphia under these rules.

It should be noted that although the allocation with complete information always dominates for a given policy goal, the ranking of the effectiveness of the four sub-optimal allocation rules, *i.e.* Rules 2-4, depends on the pairs of MSA's under consideration. As shown in Appendix Table A8, where we repeat the exercise for four other pairs of MSA's, we find that none of the four sub-optimal rules is always better than the other three. This is not surprising because the effectiveness of a sub-optimal rule depends on how close its criterion is to a sufficient statistic for the observed

and unobserved features of the MSA's. For example, in the case of Philadelphia versus Phoenix, the observable differences between the pair line up with almost all their unobservable differences in terms of the characteristic-crime relationship. Of course, without knowing the unobservables, it is impossible to tell *ex ante* which sub-optimal criterion is closer to being a sufficient statistic. Therefore, there is no "golden rule" in terms of the allocation of police when information about unobservables is not available.

7. CONCLUSION

We have developed and estimated a model of crime in which the number of police, the crime rate, the arrest rate, the employment rate, and the wage rate are joint outcomes of a subgame perfect Nash equilibrium. The local government chooses the size of its police force and citizens choose among work, home, and crime. MSA's differ in the effectiveness of their police force, the productivity of their economies, their citizens' values of leisure and the marginal cost of police. The model is estimated using MSA-level data from the FBI UCRs and from the CPS for 2008. The model was shown to match the data well.

The estimates of the model were used to examine the effects on crime of monitored federal transfers to local governments to increase the size of the police force. We found that the effectiveness of such programmes differs substantially across MSA's, and that knowledge about unobserved heterogeneity in MSA-specific characteristics is especially useful for the optimal allocation of police across MSA's. For example, in a pair-wise comparison of the Philadelphia and Phoenix MSA's, it was found that the optimal allocation of additional police between them would produce a reduction in the overall crime rate of 16%. On the other hand, an allocation of police based on their levels of crime or arrests would produce only a 5.6% reduction in crime. We did not pursue an analysis of a nationwide policy of allocating additional police among all MSA's jointly; although computationally burdensome, such an analysis would be feasible.

A major part of the empirical literature on crime has adopted an econometric strategy that is intended to approximate the solution of a behavioural optimizing model. This approach has been used to provide estimates of the responsiveness of crime rates to deterrence measures, such as apprehension rates, and to criminal justice resources, such as expenditures on policing. However, this approach is limited in terms of the kinds of policy analyses that can be performed. One example of this limitation is provided by the *ex ante* policy analyses that we have pursued in this article, namely efficiently allocating federal grants to augmenting police forces across MSA's. There are, of course, limitations to our approach as well that would form an agenda for future research. Among the extensions would be to allow for the choice of punishment severity, to allow for dynamics in the choice of criminal behaviour and to explicitly account for a government budget constraint.

APPENDIX

A1. FUNCTIONAL FORMS

We categorize each of the observable characteristics. There are four age groups: 16–25, 26–40, 41–50, 51–64; four education groups: less than high school, high school grads, some college, and colleges grads or above; two race groups: black/Hispanics and others; and three categories for the number of young children, 0, 1, and 2 or above. The coefficient λ for one group for each characteristic is restricted to be zero. For the human capital type:

$$x' \lambda^l = \lambda_0^l + \sum_{a=1}^4 \lambda_{1a}^l I(\text{age}_i = a) + \sum_{s=1}^4 \lambda_{2s}^l I(\text{edu}_i = s) + \lambda_3^l I(\text{female}) + \lambda_4^l I(\text{black/hispanic}).$$

For the taste for staying at home:

$$\begin{aligned} x' \lambda^\eta &= \lambda_0^\eta + \sum_{a=1}^4 \lambda_{1a}^\eta I(\text{age}_i = a) + \sum_{s=1}^4 \lambda_{2s}^\eta I(\text{edu}_i = s) + \lambda_3^\eta I(\text{female}) + \lambda_4^\eta I(\text{black/hispanic}) \\ &+ \sum_{k=0}^2 \lambda_{5k}^\eta I(\text{kid}_i = k) + \sum_{k=0}^2 \lambda_{6k}^\eta I(\text{kid}_i = k, \text{female}) + \sum_{s=1}^2 \lambda_{7s}^\eta I(\text{edu}_i = s, \text{age}_i = 1). \end{aligned}$$

For the taste for crime:

$$\begin{aligned} x' \lambda^\kappa &= \lambda_0^\kappa + \sum_{a=1}^4 \lambda_{1a}^\kappa I(\text{age}_i = a) + \sum_{s=1}^4 \lambda_{2s}^\kappa I(\text{edu}_i = s) + \lambda_3^\kappa I(\text{female}) + \lambda_4^\kappa I(\text{black/hispanic}) \\ &+ \sum_{k=0}^2 \lambda_{5k}^\kappa I(\text{kid}_i = k). \end{aligned}$$

A2. CROSS-MSA DIFFERENCES: IMPLEMENTATION DETAILS

Step 1: For each pair of MSA's, we draw a vector of unobservable characteristics ($u^s = (v_j, \tau_j, \epsilon_j, \bar{\kappa}_j)$) and simulate equilibrium outcomes in both MSA's using the same u^s . Let E_j^s be the set of equilibria for MSA j under u^s . A vector u^s is acceptable if, for each of the two MSA's, at least one equilibrium in E_j^s is close to the data outcomes in j . Closeness is measured by the sup norm. Specifically, letting y_{ji} be the i^{th} outcome for MSA j in the data (number of police, employment rate, crime rate, arrest rate, and average income) and y_{jie}^s the simulated outcome under Equilibrium $e \in E_j^s$, the vector s is acceptable for the MSA pair (k, k') if

$$\min_e \left\{ \max_i \left\{ \left| \frac{y_{jie}^s - y_{ji}}{y_{ji}} \right| \right\} \right\}_{e \in E_j^s} < \varepsilon \text{ for } j = k, k'.$$

Under an acceptable u^s , the equilibrium $e_j^s \in E_j^s$ that is closest to the data is called the baseline. We repeat this process to obtain a large number of acceptable vectors $\{u^s\}_s$ for each pair. We compute the average baseline outcomes across acceptable $\{u^s\}_s$. The resulting crime rates are the baseline μ_j^{BASE} . Note that a given u^s is associated with multiple equilibria for MSA j , which means the observed outcomes may arise from different (u^s, e_j^s) combinations. Therefore, we need to integrate over a large number of acceptable vectors $\{u^s\}_s$.

Step 2: for each MSA paired with Philadelphia, we focus on the "same" equilibrium as the one played in Philadelphia. For each MSA j , we select from E_j^s the equilibrium that is the rank-wise counterpart of the baseline Philadelphia equilibrium under u^s . Then, we compute the average outcome for j under the newly imposed equilibrium across acceptable u^s . The resulting crime rate is denoted as μ_j^{PHL} .

A3. MODEL DETAILS

A3.1 Lower bound of the taste for crime η^*

$$\begin{aligned} \max\{V^1|\eta\} &= \max\{(1-\mu)[(1-\pi)E(\ln(\alpha y + b)) + \pi \ln(c)] + \mu \ln(b) + \eta\} \\ &= \ln[\alpha l_{N_i} r + r l_1] + \eta \\ &= \ln(\alpha l_{N_i} + l_1) + \ln(r) + \eta \\ \min\{V^0|\eta\} &= \min\{\max\{\ln(y) + \mu \ln(1-\alpha), \ln(b) + \mu \ln(1-\alpha) + \kappa\}\} \\ &= \min\{\max\{\ln(y), \ln(b) + \kappa\} + \mu \ln(1-\alpha)\} \\ &= \min\{\max\{\ln(y), \ln(b) + \kappa\} + \ln(1-\alpha)\} \\ &\geq \ln(b) + \kappa_1 + \ln(1-\alpha) \\ &= \ln(l_1) + \ln(r) + \kappa_1 + \ln(1-\alpha) \\ \eta^* &\geq -\ln(\alpha l_{N_i} + l_1) + \ln(l_1) + \kappa_1 + \ln(1-\alpha) \end{aligned}$$

One will never be a criminal for any $\eta \leq \eta^*$.

A3.2 Number of criminal propensity types

$$\begin{aligned}
 N^* &= N_\eta \left(\sum_{j=1}^{N_\kappa} I(\ln(l_1) + \kappa_j > \min\{\ln(l)\}) \right) \\
 &\quad + N_\eta \left(\sum_{i=1}^{N_l} I(\ln(l_i) \geq \ln(l_1) + \min\{\kappa\}) \right) \\
 &= N_\eta \times N_\kappa + N_\eta \left(\sum_{i=1}^{N_l} I(\ln(l_i) \geq \ln(l_1) + \kappa_1) \right) \\
 &\leq N_\eta \times (N_\kappa + N_l - 1).
 \end{aligned}
 \tag{14}$$

The first term in (14) counts the total number of types who choose not to be employed. Conditional on declining a job offer, these individuals' decisions will not vary with l . Any (η_n, κ_n) may fall into this group as long as the first indicator function holds, which is always true given that κ is always positive, hence $N_\eta \times N_\kappa$. The second term counts the number of types who will choose between work and crime and hence whose decisions will not vary with κ . The condition in the indicator function excludes types who will choose home over work and hence are already included in the first term.

A3.3 Measure of criminal propensity types

$$h_{jn} = \begin{cases} p_{\kappa_n} p_{\eta_n} \sum_{i=1}^{N_l} p_{l_i} I(\ln(l_i) < \ln(l_1) + \kappa_n) & \text{if } T_n = \ln(l_1) + \kappa_n - \eta_n, \\ p_{\eta_n} p_{l_n} \sum_{i=1}^{N_\kappa} p_{\kappa_i} I(\ln(l_n) \geq \ln(l_1) + \kappa_i) & \text{if } T_n = \ln(l_n) - \eta_n. \end{cases}$$

The first line is the measure of individuals with (η_n, κ_n) who declined a job offer. The second line is the measure of those with (η_n, l_n) who prefer employment over non-employment.

A3.4 Unique solution to the government problem

Proof. Let $L(s_j)$ denote the government cost evaluated at s_j , *i.e.*

$$L(s_j) = v_j s_j + \sum_{n=1}^{N^*} q_j^n(s_j) \left(\omega_1 \exp(\mu_j^n) - \omega_2 \ln(\pi_j^n(s_j)) \right).
 \tag{15}$$

The proof of uniqueness is via contradiction. Suppose there exist s^* and s' , both of which minimize the government's cost. Without loss of generality, suppose $s' < s^*$. $L(s^*) = L(s')$ implies

$$v_j (s^* - s') = \omega_1 \sum_{n=1}^{N^*} \exp(\mu_j^n) (q_j^n(s^*) - q_j^n(s')) - \omega_2 \sum_{n=1}^{N^*} [q_j^n(s^*) \ln(\pi_j^n(s^*)) - q_j^n(s') \ln(\pi_j^n(s'))].$$

The left hand side (LHS) $v_j (s^* - s') > 0$ because $v_j > 0$ and $s' < s^*$.

The right hand side (RHS) can be written as

$$\begin{aligned}
 &\omega_1 \left(\sum_{n=1}^{N^*} \exp(\mu_j^n) q_j^n(s^*) - \sum_{n=1}^{N^*} \exp(\mu_j^n) q_j^n(s') \right) \\
 &- \omega_2 \sum_{n=1}^{N^*} (q_j^n(s^*) - q_j^n(s')) \ln(\pi_j^n(s^*)) \\
 &+ \omega_2 \sum_{n=1}^{N^*} q_j^n(s') (\ln(\pi_j^n(s')) - \ln(\pi_j^n(s^*))).
 \end{aligned}
 \tag{16}$$

We have the following observations

- (1). The first line is non-positive, because a larger police force will not change the distribution in equilibria in a way that increases the expected loss from crime, otherwise, s^* cannot be cost-minimizing.
- (2). Notice that in the second line, $\pi_j^n(s^*)$ is the arrest rate given s^* and the crime rate μ_j^n . The second line is non-positive for the same reason as in (1): a larger police force will not change the distribution in equilibria in a way that puts more weights on higher-crime (lower-arrest) equilibria.
- (3). The third line is negative: $\pi_j^n(s') < \pi_j^n(s^*)$ for any n , because they are both under the n^{th} possible crime rate, *i.e.*, the same μ_j^n , and the arrest rate increases with the size of the police force.

In sum, we have the LHS > 0, and RHS < 0: a contradiction. ||

A4. A MORE FLEXIBLE ARREST RATE FUNCTION:

$$\pi_j = \Pi(s_j, \mu_j, \epsilon_j) = \exp\left(-\frac{\gamma_j(\mu_j \epsilon_j)^\rho}{s_j}\right), \quad (17)$$

where $\rho > 0$, $\epsilon_j \sim \ln N(-0.5\sigma_\epsilon^2, \sigma_\epsilon^2)$, and γ_j is given by

$$\begin{aligned} \gamma_j = & b_0 + b_1 \ln(1 + P_j(\text{edu} = 1)) + b_2 \ln(1 + P_j(\text{age} = 1)) \\ & + b_3 \ln(1 + P_j(\text{male})) + b_4 \ln(1 + P_j(\text{black})) + b_5 \ln(1 + P_j(\text{kid} = 0)). \end{aligned}$$

The counterfactual results are robust to the more flexible specifications, as in Table A19.

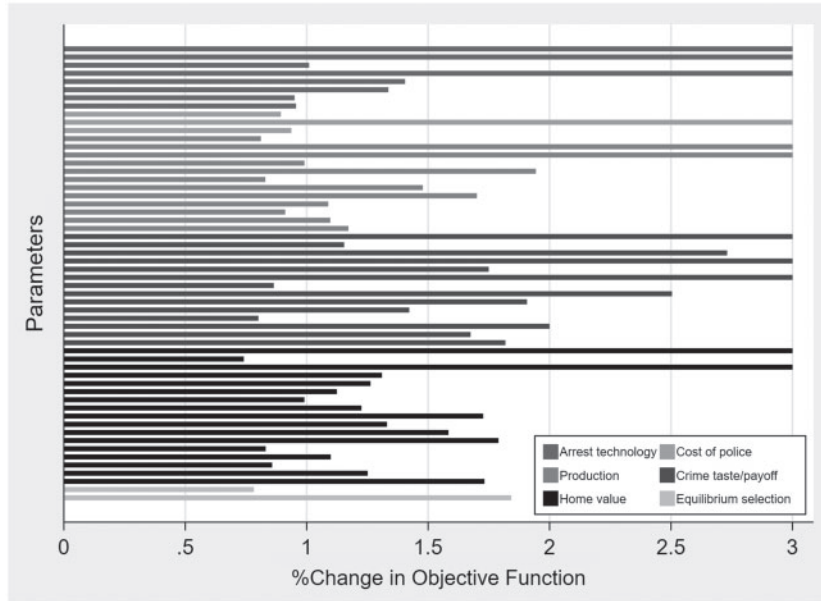


FIGURE 1

A5. SENSITIVITY OF THE OBJECTIVE FUNCTION WITH RESPECT TO PARAMETERS

Figure 1 shows the sensitivity of the objective function value with respect to each of the parameters of the model. For each structural relationship, the arrest technology, the production function, etc., the figure plots the percentage change in the value of the objective function following a given percentage change in a parameter from its estimated value. As seen, the changes in the objective function ranges from 0.8% to over 10% (the figure is truncated to the right). As argued in [Adda *et al.* \(2017\)](#), the magnitude of these changes in the objective function to discrete value changes in the parameters indicate that the objective function is not flat. This evidence is consistent with the great majority of the parameters having relatively small standard errors.

A6. IMPOSING EQUILIBRIUM CONSISTENCY ACROSS YEARS WITHIN AN MSA

We re-estimate the model using both 2003 and 2008 data, with the restriction that the same ranked equilibrium was chosen in both years within the same MSA. Only 181 MSA's have matched CPS and FBI data in both years. The new model is estimated using these 181 MSA's in both years, while the original model estimation uses 238 MSA's in 2008. Table A20 shows selected parameter estimates from this restricted model (Column 3) along with the original estimates (Column 2). Most estimates are similar across the two columns. With the additional restriction in the equilibrium selection rule, the

TABLE A1
Response to 1 std dev. increase in each MSA-level shock

%	Productivity(τ)		Arrest efficiency (ϵ)		Leisure value ($\bar{\kappa}$)		Police cost (ν)
	Gov. Move ^a	Police Fixed ^b	Gov. Move	Pol. Fixed	Gov. Move	Pol. Fixed	Gov. Move
Police	-0.43	0	-13.96	0	-0.41	0	14.83
Crime	-2.81	-6.30	-11.91	-17.73	-3.61	-8.27	-7.83
Arrest	0.37	1.43	35.36	61.47	-0.02	2.07	23.54

^aLocal governments adjust police size.

^bPolice size fixed at baseline.

TABLE A2
Home taste: λ^k and σ_k

Less than HS	0.04	(0.02)
Some College	0.10	(0.03)
BA and Above	0.38	(0.02)
Age<26	0.14	(0.03)
41<Age<51	0.70	(0.03)
Age>50	1.03	(0.04)
Female	0.20	(0.03)
Black/Hispanic	-0.03	(0.02)
1 Kid	-0.98	(0.08)
Multiple Kids	-1.14	(0.06)
1 Kid, Female	1.73	(0.05)
Multiple Kids, Female	2.38	(0.07)
Age<26, Less than HS	0.95	(0.04)
Age<26, HS	-3.06	(0.26)
Constant	-1.74	(0.01)
σ_k	0.85	(0.01)

Default group: HS grads, aged 26–40, male, non-black/Hispanic, and no kid.

TABLE A3
Human capital: λ^l and σ_l

Less than HS	-0.56	(0.02)
Some College	0.19	(0.01)
BA and Above	0.84	(0.02)
Age<26	-0.72	(0.02)
41<Age<51	0.25	(0.01)
Age>50	0.09	(0.01)
Female	-0.53	(0.02)
Black/Hispanic	-0.10	(0.02)
Constant	0.87	(0.02)
σ_l	1.05	(0.01)

Default group: HS grads, aged 26–40, male, and non-black/Hispanic.

model finds slightly higher dispersions of most MSA-level unobservables. In addition, both ζ_1 and ζ_2 become slightly more negative, meaning that the distribution of equilibria is more concentrated around the middle-ranked equilibrium.

Table A21 shows the fit under the restricted model, where the middle and lower panels show the fit of within-MSA cross-year outcome differences. Table A22 contrasts the counterfactual experiment results from the original model, and those from the new restricted model, where there is a 20% increase in police force in all MSA's. In both cases, we use the same 181 MSA's. The new model predicts a lower elasticity of crime to police force at -0.29 , while the original model predicts it at -0.40 . The latter is closer to the -0.38 figure in [Draca et al. \(2011\)](#).

TABLE A4
Crime taste: λ^η and σ_η

Less than HS	0.47	(0.01)
HS Grads	0.55	(0.01)
Some college	-0.04	(0.04)
Age < 26	0.35	(0.02)
41 < Age < 51	-1.19	(0.10)
Age > 50	-0.36	(0.06)
Female	-0.18	(0.01)
Black/Hispanic	0.16	(0.02)
1 Kid	-2.88	(0.11)
Multiple Kids	-3.63	(0.20)
Constant	-1.32	(0.01)
σ_η	0.69	(0.01)

Default group: BA and above, aged 26–40, male, non-black/Hispanic, and no kid.

TABLE A5
Pr(school enrolment | age < 26, edu, home)

Less than HS	0.97	(0.04)
HS	0.92	(0.05)
Some College	0.96	(0.15)
BA and above	0.44	(0.04)

Problem of enrolment by current education for youth who choose the home option.

TABLE A6
Gov. preference and arrest

Gov. weight on crime ω_1	28.71	(1.21)
Gov. weight on arrest ω_2	2.03	(0.03)
Arrest technology b_0	21.32	(0.23)
Arrest technology b_1	35.05	(2.09)
Arrest technology ρ	0.06	(0.001)

TABLE A7
National outcomes: uniform 20% increase in police

	Original Model	High-Crime-Only	Low-Crime-Only
% Change in Crime Rate	-8.23	-12.21	-1.86
Elasticity Crime to Police	-0.41	-0.61	-0.09
<u>% Reduction in Crime Rate</u> Additional police per 1000 people	8.80	12.86	1.96

TABLE A8
Police allocation rules and crime reduction

%	Pair 1	Pair 2	Pair 3	Pair 4
Complete info	15.20	8.42	14.70	16.01
Based on x	5.51	2.76	13.51	14.62
Based on crime rates	5.51	6.32	5.58	11.90
Based on arrest rates	5.51	6.32	5.58	5.37
Based on GDP	11.46	6.32	11.18	7.86

Pair 1: Atlanta-Sandy Springs-Marietta versus Philadelphia

Pair 2: Dallas-Plano-Irving versus Phoenix-Mesa-Scottsdale

Pair 3: Fort Lauderdale-Pompano Beach-Deerfield Beach versus Philadelphia

Pair 4: Philadelphia versus Washington-Arlington-Alexandria

TABLE A9
Police allocation and crime reduction: Philadelphia versus Phoenix (Most Likely Crime)

	% Share of extra police for Philadelphia	% Reduction*
Complete information	87.0	16.00
Based on x	96.0	15.07
Based on crime rates	47.2	5.61
Based on arrest rates	40.8	5.61
Based on GDP	51.4	5.61

*Cross-MSA average crime rate from the most likely equilibrium.

TABLE A10
Police allocation rules and crime reduction (Most Likely Crime)

%	Pair 1	Pair 2	Pair 3	Pair 4
Complete info	15.30	10.60	15.97	15.86
Based on x	5.56	8.47	14.88	14.53
Based on crime rates	7.84	8.47	13.68	8.06
Based on arrest rates	5.56	8.47	7.81	5.67
Based on GDP	14.04	8.47	13.68	14.53

Pair 1: Atlanta-Sandy Springs-Marietta versus Philadelphia
 Pair 2: Dallas-Plano-Irving versus Phoenix-Mesa-Scottsdale
 Pair 3: Fort Lauderdale-Pompano Beach-Deerfield Beach versus Philadelphia
 Pair 4: Philadelphia versus Washington-Arlington-Alexandria

TABLE A11
Compare stats: agencies with and without missing reports

	Agencies w/o missing reports		Agencies with missing reports	
	Mean	Std. dev.	Mean	Std. dev.
Population (1,000)	23.4	109.5	4.8	13.2
Police/population* (%)	0.33	1.1	0.33	0.5
Num. of obs.	14,920		6,433	

*At the agency level.

TABLE A12
Compare stats: matched versus unmatched MSA's

	MSA in sample		Unmatched MSA ^a	
	Mean	Std. dev.	Mean	Std. dev.
Population (1,000)	588.4	1195.8	142.7	296.8
Crime (%)	5.65	1.70	5.01	1.59
Arrest (%)	19.03	6.85	20.4	7.29
Police/population ^b (%)	0.48	0.19	0.43	0.13
Num. of obs.	238		121	

^aMSA's in the UCR data that failed to match the CPS data.

^bAt the MSA level.

TABLE A13
Outcomes by individual characteristics

	Employment rate (%)		Earnings working (\$)	
	Data	Model	Data	Model
Less than HS	47.0	47.4	19,199	18,626
HS	73.8	74.3	31,128	30,110
Some College	77.0	77.7	35,398	33,914
BA and Above	85.2	85.8	65,245	62,635
Age<26	54.1	53.5	18,502	18,574
41<Age<51	82.3	82.8	49,908	48,789
Age>50	71.4	72.4	49,951	47,848
Female	67.6	67.1	34,203	34,128
Black/hispanic	68.9	68.6	31,454	31,969
1 Kid	75.1	76.5	44,936	41,552
Multiple kids	71.1	72.0	48,430	44,259
1 Kid, female	62.8	63.0	32,999	33,522
Multiple kids, female	53.7	52.9	31,064	36,922

TABLE A14
Model fit

% MSA's with crime rates below n th percentile									
Data	10	20	30	40	50	60	70	80	90
Model	8.4	15.3	25.8	38.2	49.7	62.5	70.0	82.9	92.1
Crime rate in MSA's with % of low education below n th percentile									
Data	4.9	5.0	5.1	5.2	5.2	5.3	5.4	5.5	5.5
Model	4.5	4.6	4.9	5.0	5.1	5.2	5.3	5.4	5.5
Crime rate in MSA's with % of high education below n th percentile									
Data	6.0	6.0	5.9	5.9	5.8	5.8	5.8	5.8	5.7
Model	6.4	6.3	6.2	6.0	5.9	5.9	5.8	5.8	5.7

TABLE A15
Model fit

(%)	Police		Arrest rate	
	Data	Model	Data	Model
% of low edu < 25th percentile	4.36	4.53	19.7	20.3
% of low edu < 50th percentile	4.37	4.59	19.9	19.7
% of low edu < 75th percentile	4.61	4.62	19.4	19.4
% of age below 26 < median	4.87	4.63	18.9	18.8
% of age above 50 < median	4.65	4.77	17.8	19.1
% of male < median	4.89	4.71	18.9	19.2

TABLE A16
Means (2003)

	Crime (%)	Arrest (%)	Police (%)	Employment (%)	Mean wage (\$)	Std dev wage (\$)
Data	6.32	17.00	0.47	74.40	39,270	45,190
Model	5.98	18.20	0.48	75.38	38,434	39,265

TABLE A17
Standard deviations (2003)

	Crime (%)	Arrest (%)	Police (%)	Employment (%)	Mean wage (\$)
Data	1.81	6.19	0.17	5.60	6,583
Model	1.32	7.81	0.13	3.12	5,328

TABLE A18
Correlations (2003)

	Data	Model
(Police, crime)	0.28	0.14
(Police, arrest)	-0.03	-0.02
(Police, employment)	-0.14	-0.09
(Police, average wage)	0.02	-0.04
(Employment, crime)	-0.30	-0.34
(Employment, arrest)	0.13	0.23
(Employment, average wage)	0.12	0.19
(Crime, arrest)	-0.33	-0.48
(Crime, average wage)	-0.23	-0.28
(Arrest, average wage)	-0.15	0.03

TABLE A19
National outcomes: uniform 20% increase in police

	Original model	More flexible arrest function
% Change in crime rate	-8.23	-8.61
Elasticity crime to police	-0.41	-0.43
$\frac{\% \text{ reduction in crime rate}}{\text{Federal-funded additional police per 1,000 people}}$	8.80	8.98

TABLE A20
Selected parameter estimates

Sample	Original estimates	Restricted model
	238 MSAs, Year 2008	181 MSAs, Years 2003 and 2008
Production θ	0.84	0.84
Return to crime α	0.10	0.10
Consumption if arrested \underline{c}	0.09	0.08
Dispersion of MSA-level unobservables		
Productivity σ_{τ}	0.12	0.07
Arrest efficiency σ_{ϵ}	5.89	6.11
Value of leisure $\sigma_{\bar{c}}$	0.13	0.52
Marginal cost of police σ_v	0.25	0.33
Probability distribution of market equilibria		
Below median slope ζ_1	-0.20	-0.39
Above median slope ζ_2	-0.60	-0.65

TABLE A21
Fit: restricted model

	Crime (%)	Arrest (%)	Police (%)	Employment (%)	Mean wage (\$)
Mean (y)					
Data	6.11	17.78	0.48	73.80	39,177
Model	6.08	17.72	0.48	74.93	40,600
Mean ($y_{2008} - y_{2003}$)					
Data	-0.78	1.50	0.02	-1.21	-675
Model	-0.17	1.03	0.01	0.44	448
Std dev ($y_{2008} - y_{2003}$)					
Data	1.18	4.72	0.11	6.33	7,283
Model	1.27	3.76	0.07	3.48	2,969

TABLE A22
National outcomes: uniform 20% increase in police

	Original model	Restricted model
% Change in crime rate	-8.10	-5.78
Elasticity crime to police	-0.401	-0.29
Sample: 181 MSA, Year 2008		

Acknowledgments. We thank Fumihiko Suga and Xiaodong Fan for excellent research assistance. We thank Steven Durlauf, John Kennan, Thomas MaCurdy, Xiaoxia Shi, and Chris Taber for insightful discussions and workshop participants at SITE conference, Ohio State University, St. Louis FED, University of Virginia for helpful comments. We thank the editor and four anonymous referees for their inputs.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- ADDA, J., McCONNELL, B. and RASUL, I. (2014), "Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment", *Journal of Political Economy*, **122**, 1130–1202.
- ADDA, J., DUSTMANN, C. and STEVENS, K. (2017), "The Career Costs of Children", *Journal of Political Economy*, **125**, 293–337.
- ACKERBERG, D. and GOWRISANKARAN, G. (2006), "Quantifying Equilibrium Network Externalities in the ACH Banking Industry", *The RAND Journal of Economics*, **27**, 738–761.
- ANDERSON, M. (2014), "In School and Out of Trouble? The Minimum Dropout Age and Juvenile Crime", *Review of Economics and Statistics*, **96**, 318–31.
- ANGRIST, J. and IMBENS, G. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, **62**, 467–475.
- BAJARI, P., HONG, H. and RYAN, S. (2010), "Identification and Estimation of a Discrete Game of Complete Information", *Econometrica*, **78**, 1529–1568.
- BECKER, G. (1968), "Crime and Punishment: An Economic Approach", *The Journal of Political Economy*, **76**, 169–217.
- BJORN, P. and VUONG, Q. (1984), "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with Application to Labor Force Participation" (SSWP 537, California Institute of Technology).
- BURDETT, K., LAGOS, R. and WRIGHT, R. (2003), "Crime, Inequality, and Unemployment", *American Economic Review*, **93**, 1764–1777.
- BURDETT, K., LAGOS, R. and WRIGHT, R. (2004), "An On-the-Job Search Model of Unemployment, Inequality, and Crime", *International Economic Review*, **45**, 681–706.
- CARD, D. and GIULIANO, L. (2013), "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends", *Review of Economic and Statistics*, **95**, 1130–1149.
- CLAY, K., LINGWALL, J. and STEPHENS, M. (2012), "Do Schooling Laws Matter? Evidence from the Introduction of Compulsory Attendance Laws in the United States" (NBER Working Paper 18477).

- CHALFIN, A. and McCrARY, J. (2013), "Are U.S. Cities Underpoliced? Theory and Evidence" (NBER Working Paper #18815).
- CHALFIN, A. and McCrARY, J. (2017), "Criminal Deterrence: A Review of the Literature" *Journal of Economic Literature*, **55**, 5–48.
- CONLEY, J. and WANG, P. (2006), "Crime and Ethics", *Journal of Urban Economics*, **60**, 107–123.
- COOK, P. and KANG, S. (2016), "Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation", *American Economic Journal: Applied Economics*, **8**, 33–57.
- DRACA, M. and MACHIN, S. (2015), "Crime and Economic Incentives" (Manuscript, UCL).
- DRACA, M., MACHIN, S. and WITT, R. (2011), "Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks." *American Economic Review*, **101**, 2157–2181.
- DURLAUF, S., NAVARRO, S. and RIVERS, D. (2010), "Understanding Aggregate Crime Regressions", *Journal of Econometrics*, **158**, 306–317.
- EHRlich, I. (1973), "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation", *The Journal of Political Economy*, **81**, 521–565.
- EHRlich, I. (2010), "The Market Model of Crime: A Short Review and New Directions", in Benson, B. L. and Zimmerman, P. R., eds, *The Handbook on the Economics of Crime* (Northampton, MA: Edward Elgar) 3–23.
- ENGELHARDT, B., ROCHETEAU, G. and RUPERT, P. (2008), "Crime and the Labor Market: A Search Model with Optimal Contracts", *Journal of Public Economics*, **92**, 1876–1891.
- EVANS, W. and OWENS, E. (2007), "COPS and Crime", *Journal of Public Economics*, **91**, 181–201.
- FELLA, G. and GALLIPOLI, G. (2014), "Education and Crime Over the Life Cycle", *Review of Economic Studies*, **81**, 1484–1517.
- FENDER, J. (1999), "A General Equilibrium Model of Crime and Punishment", *Journal of Economic Behavior & Organization*, **39**, 437–453.
- FLINN, C. (1986), "Dynamic Models of Criminal Careers", in Blumstein, A. et al., eds, *Criminal Careers and "Career Criminals"* (Washington, DC: National Academy Press) 356–379.
- FREEMAN, S., GROGGER, J. and SONSTELIE, J. (1996), "The Spatial Concentration of Crime", *Journal of Urban Economics*, **40**, 216–231.
- GLAESER, E., SACERDOTE, B. and SCHEINKMAN, J. (1996), "Crime and Social Interactions", *Quarterly Journal of Economics*, **111**, 507–548.
- GROGGER, J. (1998), "Market Wages and Youth Crime", *Journal of Labor Economics*, **16**, 756–791.
- HECKMAN, J. (1971), "Three Essays on the Supply of Labor and the Demand for Goods", (unpublished Ph.D. dissertation, Department of Economics, Princeton University, Princeton, N.J.).
- HECKMAN, J. and VYTLACIL, E. (1998), "Instrumental Variables Methods for the Correlated Random Coefficient Model", *Journal of Human Resources*, **33**, 974–1002.
- HJALMARSSON, R., HOLMLUND, H. and LINDQUIST, M. (2015), "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-Data", *Economic Journal*, **125**, 1290–1326.
- HUANG, C., LIANG, D. and WANG, P. (2004), "Crime and Poverty: A Search-theoretic Approach", *International Economic Review*, **45**, 909–938.
- JIA, P. (2008), "What Happens When Wal-Mart Comes to Town; An Empirical Analysis of the Retail Discounting Industry", *Econometrica*, **76**, 1236–1316.
- IMAI, S. and KRISHNA, K. (2004), "Employment, Deterrence, and Crime In A Dynamic Model", *International Economic Review*, **45**, 845–872.
- IMROHORGLU, A., MERLO, A. and RUPERT, P. (2000), "On the Political Economy of Income Redistribution and Crime", *International Economic Review*, **41**, 1–25.
- IMROHORGLU, A., MERLO, A. and RUPERT, P. (2004), "What Accounts for the Decline in Crime", *International Economic Review*, **45**, 707–729.
- LEVITT, S. and LOCHNER, L. (2000), "The Determinants of Juvenile Crime", in Gruber, J., ed., *Risky Behavior among Youths: An Economic Analysis* (Chicago, IL: University of Chicago Press).
- LEVITT, S. (2002), "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply." *American Economic Review*, **92**, 1244–1250.
- LIN, M. (2009), "More Police, Less Crime: Evidence from U.S. State Data", *International Review of Law and Economics*, **29**, 73–80.
- LOCHNER, L. and MORETTI, E. (2004), "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports", *American Economic Review*, **94**, 155–189.
- MACHIN, S., MARIE, O. and VUJIC, S. (2011), "The Crime Reducing Effect of Education", *Economic Journal*, **121**, 463–84.
- PARUCHURI, P., PEARCE, J., MARECKI, J., et al. (2008), "Efficient Algorithms to solve Bayesian Stackelberg Games for Security Applications", in AAA'08: *Proceedings of the 23rd National Conference on Artificial Intelligence*, Vol. 3 (AAAI Press) 1559–1562.
- SAH, R. (1991), "Social Osmosis and Patterns of Crime", *Journal of Political Economy*, **99**, 1272–1295.
- TAUCHEN, H. (2010), "Estimating the Supply of Crime: Recent Advances", in Benson, B. L. and Zimmerman, P. R., eds, *The Handbook on the Economics of Crime* (Northampton, MA: Edward Elgar) 24–52.
- VERDIER, T. and ZENOU, Y. (2004), "Racial Beliefs, Location, and the Causes of Crime", *International Economic Review*, **45**, 731–760.

- WORRALL, J. and KOVANDZIC, T. (2010), "Police Levels and Crime Rates: An Instrumental Variables Approach", *Social Science Research*, **39**, 506–516.
- ZHANG, C., JAIN, M., GOYAL, R., *et al.* (2015), "Keeping Pace with Criminals: Learning, Predicting and Planning against Crime: Demonstration Based on Real Urban Crime Data (Demonstration)", International Conference on Autonomous Agents and Multiagent Systems (AAMAS).
- ZHAO, J., SCHEIDER, M. C. and THURMAN, Q. (2002), "Funding Community Policing to Reduce Crime: have COPS Grants made a Difference?" *Criminology and Public Policy*, **2**, 7–32.
- ZIPKIN, J., SHORT, M. and BERTOZZI, A. (2014), "Cops on the Dots in a Mathematical Model of Urban Crime and Police Response", *Discrete and Continuous Dynamical Systems B*, **19**, 1479–1506.