Education and Crime over the Life Cycle

Giulio Fella*

Giovanni Gallipoli†

Abstract

We develop an overlapping-generation, life-cycle model with endogenous education and crime choices. Education and crime depend on different dimensions of heterogeneity. We apply the model to property crime and calibrate it to U.S. data. We compare two policies: subsidizing high school completion and increasing the length of prison sentences. We find that targeting crime reductions through increases in high school graduation rates entails large efficiency and welfare gains. These gains are absent if the same crime reduction is achieved by increasing the length of sentences. We find that general equilibrium effects explain roughly one half of the reduction in crime from subsidizing high school. Crucially, the effect of small equilibrium price changes is magnified by their interaction with the underlying individual heterogeneity.

JEL classification: H52

Keywords: Crime, education, equilibrium, policy.

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. E-mail: `g.fella@qmul.ac.uk`

†Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail: `gallipol@interchange.ubc.ca`

1 Introduction

Crime is a hot issue on the U.S. policy agenda. Despite its significant fall in the Nineties its cost to the taxpayer has soared. The prison population has doubled over the same period and now stands at over two millions of inmates. The average annual cost per prison inmate was 28,900 dollars in 2008.[1] The dramatic increase in the U.S. prison population over the past twenty years has prompted a shift of interest, among both academics and policymakers, from tougher sentencing to other forms of intervention.[2]

This paper develops a heterogeneous-agent, equilibrium life-cycle model incorporating both education and criminal choices. Its goal is to provide a framework within which to compare the effectiveness and, importantly, the welfare implications of alternative policies which directly or indirectly impact on crime. The model emphasizes the crucial role of heterogeneity. Agents differ in: 1) innate, observed ability, 2) innate crime propensity[3] and 3) initial wealth. Agents self-select into education on the basis of these differences and, upon entering the labor market, decide whether to engage in criminal activity on a period-by-period basis.

We apply the model to the study of property crime which - unlike, for example, violent crime - is most likely to be driven by economic considerations. We calibrate the model to data for the US economy in 1980 and verify that its predictions are consistent with a number of untargeted features of the data.[4] The model is also tested by assessing its ability to account for changes in the victimization rate over time (between 1980 and 2000).

We use the model to evaluate two alternative policies: a subsidy towards high school completion and an increase in the prison sentence which generates the same change in the victimization rate as the subsidy policy. The size of the subsidy is 8.8% per cent of

[1]Source: Pew Center on the States, "One in 31: The Long Reach of American Corrections" (Washington, DC: The Pew Charitable Trusts, March 2009).

[2]See, e.g., the survey by Donohue and Siegelman (2004).

[3]Ability is a set of characteristics that directly affects earnings, whereas crime propensity residually captures in a single index the range of unmodelled individual, family and social characteristics that contribute to shaping individual attitudes towards crime (e.g. gender, broken families, neighborhood quality). Merlo and Wolpin (2009) provide evidence supporting the importance of residual heterogeneity in accounting for criminal behavior.

[4]This is mostly, but not exclusively, done by running identical estimation procedures on simulated data from our benchmark economy as well as comparable data.

average labor earnings per year of schooling and is chosen to coincide with the average value of the monetary component in a well-known, small scale program.[5]

We consider the effect of making the subsidy available to everybody completing high school. The policy increases the equilibrium share of high school graduates in the population by one percentage point and reduces the victimization rate by roughly 8 per cent relative to the benchmark. Even more importantly, the subsidy implies significant efficiency and welfare gains. Efficiency, as measured by the steady state flow of aggregate consumption, increases by 1.3 per cent. Welfare, as measured by the permanent consumption equivalent of ex ante, expected lifetime utility, increases by over 3 per cent. This basic finding survives a variety of robustness checks and alternative parameterizations.

Compared to an unconditional high school subsidy, an increase in the prison term that induces the same fall in the victimization rate generates no efficiency or welfare gains. Intuitively, the efficiency gains of the subsidy come from its effect on the education composition of the labor force. No such effect is present in the case of the prison term. Concerning welfare, the only effect of the increase in the prison term is to reduce the transitory income risk associated with being the victim of a crime, but this is offset by the cost of financing the increased prison expenditure.[6] On the other hand, by increasing the relative price of labor for high school dropouts and weakening the link between wealth at birth and selection into education, the subsidy provides insurance against the (ex-ante) uncertain over ability and initial wealth draws. Since the ability shock is permanent and the initial wealth shock has a persistent effect through the education choice, the welfare benefits are large as they cumulate over the whole lifetime.

Conducting the same subsidy experiment in partial equilibrium reveals that the general equilibrium increase in the relative price of high school dropouts, although small, plays an important role in terms of crime reduction. Even small changes in prices have significant effects through their interaction with the underlying heterogeneity. The crime fall associated with the subsidy is only half as large as in general equilibrium.

The model is in the tradition of economic models of crime which goes back to Becker's (1968) seminal work. It builds upon the original contribution of Imrohoroglu, Merlo, and

[5]The program in question is called the Quantum Opportunities Program (QOP). See Hahn, Leavitt, and Aaron (1994) and Taggart (1995) for a discussion of the program and its effects.

[6]Note that crime is a pure redistribution and entails no deadweight loss in the model.

Rupert (2004) who are the first to construct a calibrated, structural, general equilibrium model of rational crime choice.[7] Their model is extremely successful in accounting for the evolution of the U.S. property victimization rate over the period 1980-96 on the basis of changes in wage inequality, employment opportunities, age and education distributions and expected punishment. The focus of their analysis is positive. With the aim of accounting for changes in the victimization rate, they take the education distribution as exogenous and let it vary according to its evolution in the data over the relevant period. We extend Imrohoroglu, Merlo, and Rupert's (2004) framework by endogenizing investment in education and the marginal returns to education.

There is an extensive body of empirical literature testing the main prediction of the rational theory of crime that both market returns and the expected punishment are significant determinants of criminal choices.[8] The existence of a relationship between crime and education is documented by Lochner and Moretti (2004). They find that high school graduation significantly reduces participation in both property and violent crime.

Donohue and Siegelman (2004) assess the cost-effectiveness of alternative policies aimed at tackling crime, including social policies. Their cost-benefit analysis, though, relies on elasticities from existing empirical studies.

The structure of the paper is the following. Section 2 introduces the model. Section 3 proposes a simplified analytical example to make sense of the model and summarize its main mechanisms. Section 4 discusses the numerical parameterization. Section 5 simulates the model, assesses its performance and studies the effect of alternative policies. Section 6 presents a sensitivity analysis. Section 7 concludes.

[7] Two other dynamic equilibrium models of crime are Imrohoroglu, Merlo, and Rupert (2000) and Cozzi (2006). For reasons similar to those highlighted in Heckman, Lochner, and Taber (1998), the analysis of alternative policies to tackle crime benefits from the use of a dynamic equilibrium framework. The case for the use of models allowing for equilibrium effects in policy analysis has been recently argued by various authors in different fields. See, among others, Lee and Wolpin (2004), Cunha, Heckman, and Navarro (2004) and Abbott, Gallipoli, Meghir, and Violante (2012).

[8] See, among others, Grogger (1998), and Freeman (1999), Gould, Weinberg, and Mustard (2002), Machin and Meghir (2004), Raphael and Ludwig (2003),Levitt (1997) and Levitt (1996).

2 The model

2.1 Demographics and the life cycle

Demographics: The model has an overlapping generation structure with $\bar{j}+1$ generations alive at each date. We denote by $j \in J = \{0, 1, ..., \bar{j}\}$ the age of an individual and by λ_j the conditional probability of surviving from age j to $j+1$, with $\lambda_{\bar{j}} = 0$. The unconditional probability of surviving up to age j is $\Lambda_j = \Pi_{s=0}^{j-1} \lambda_s$.[9] The size of a newborn cohort is normalized to 1.

Life cycle: Individuals go through three stages in their life cycle. In the first stage, they attend school and acquire education. An individual's educational attainment is denoted by $e \in \{L, H, C\}$, where L stands for less than high school, H for high school and C for college.

At the start of life ($j = 0$) individuals choose between dropping out of high school and entering the labor market, or studying towards a high school degree. The latter choice entails attending high school until, and including, age $j_H - 1$. At age j_H a high school graduate chooses between entering the labor market or studying towards a college degree until, and including, age $j_C - 1$. We denote by $i_e^s \in \{0, 1\}$ the choice of studying towards degree $e = H, C$. For tractability, we do not allow for flexibility in the timing of education choices and assume full commitment to the completion of an ongoing education cycle.[10] For the same reason, we do not allow individuals to work or engage in crime while in education.

The second stage of the life cycle begins when individuals start working; that is at age 0 for high school dropouts, age j_H for high school graduates and age j_C for college graduates. During the work stage, individuals can be either out of jail or in jail. Individuals out of jail supply their labor endowment inelastically, can be the victim of a theft, choose how many thefts $\tau_j \in \{0, 1, \ldots, \bar{\tau}\}$ to commit[11] and how much to consume/save. If

[9]By the law of large numbers, Λ_j is also the mass of agents of age j in the population.

[10]As there is no accrual of information while in school, a time-inconsistency problem would arise only if students could drop out without repaying the education subsidy received.

[11]According to Table 6.31 in Maguire and Pastore (1983) only 30 per cent of the state prison population in 1979 (the closest survey year to our calibration year) was unemployed at the time of incarceration. The fact that a majority of people engaged in crime are employed is also consistent with the evidence in Grogger (1998).

engaging in crime, they are apprehended and go to jail with positive probability. The consumption/saving choice of individuals out of jail takes place after all uncertainty, including that concerning apprehension, has been resolved. Also, at age j_b, an individual chooses the size of a one-off monetary transfer b to her newborn offspring.[12] Individuals in jail (convicted criminals) just consume the exogenous amount $\bar{c}$.

Finally, the last stage of the life cycle, retirement, begins at age j_r. During this stage, individuals neither work nor engage in crime. They receive a lump-sum pension *pen* from the government and choose consumption/saving over the remaining lifetime.

2.2 Preferences

Individuals have time-separable preferences and discount the future at rate ξ. The felicity function changes over the life cycle to reflect the different choices available to an individual at different ages.

In the study stage of the life cycle the felicity function takes the form

$$U^s(c_j) = u(c_j) + \mathbb{I}_{j<j_H}\psi^H(\theta) + (1 - \mathbb{I}_{j<j_H})\psi^C(\theta). \tag{1}$$

The first addendum captures the utility of current consumption c_j with the function $u(\cdot)$ being strictly increasing, concave, continuously differentiable and satisfying the Inada condition. The remaining two terms capture the utility cost of studying towards degree $e = H, C$. The disutility of education $\psi^e(\theta)$ changes with fixed individual ability θ and residually captures, in reduced form, heterogeneity in environmental and other unmodelled factors that contribute to the cross-sectional variability in educational achievement.[13] The indicator function $\mathbb{I}_{j<j_H}$ takes value one if age $j < j_H$ – the individual is attending high school – and zero otherwise.

For individuals in the work stage of the life cycle the felicity function is

$$U^w(c_j, \tau_j, b) = u(c_j) + \mathbb{I}_{\tau_j}\chi + \mathbb{I}_{j=j_b}v(b). \tag{2}$$

[12] We assume that $\lambda_j = 1$ for all $j < j_b$, so that each agent has exactly one child.

[13] For example, Heckman, Lochner, and Todd (2006) discuss the importance of (heterogeneous) psychic costs to reconcile observed enrollment rates and market returns to education.

The second addendum captures the fixed, individual-specific, utility/disutility χ of engaging in crime, with $\mathbb{I}_{\tau_j}$ an indicator function equal to one if $\tau_j > 0$ and zero otherwise. Though modeled as a utility cost, the permanent shock χ is meant to represent, in reduced form, heterogeneity in permanent characteristics which affect an individual's propensity to engage in crime above and beyond modeled economic incentives. It plays the same role as an individual fixed effect in criminal behavior and residually captures forms of heterogeneity such as gender, neighborhood, race, broken families that we do not explicitly model. For this reason, we refer to it as the crime fixed effect in what follows.[14] The third addendum, which accrues only at the bequest age j_b, is the warm-glow utility from bequeathing b to one's offspring. The function $v(\cdot)$ is strictly increasing, concave and continuously differentiable.

In the retirement stage the felicity function has the form $U^r(c_j) = u(c_j)$.

2.3 Technology and markets

Production technology: The representative firm produces output using the production function

$$Q(H,K) = H^{1-\phi}K^{\phi}, \qquad 0 < \phi < 1, \tag{3}$$

where H and K denote, respectively, the aggregate stock of human and physical capital. The human capital stock H is the aggregate

$$H = [s_L H_L^{\varrho} + s_H H_H^{\varrho} + (1 - s_L - s_H)H_C^{\varrho}]^{\frac{1}{\varrho}}. \tag{4}$$

of the stocks of human capital with education $\{L, H, C\}$. Workers with the same level of educational attainment are perfect substitutes. Physical capital depreciates at the exogenous rate δ.

Market arrangements: Markets for the four factors of production and the unique final

[14] Our modeling approach is consistent with a vast literature, especially in criminology, which identifies different patterns of criminal behavior and attributes them to differences in persistent unobserved traits ("offender types"). See, for example, Nagin and Paternoster (1991), Nagin and Land (1993), Nagin, Farrington, and Moffitt (1995) and Broidy et al. (2003). In the economic literature Merlo and Wolpin (2009) also find that unobserved heterogeneity in permanent traits ("initial conditions") plays a crucial role in explaining the crime and schooling experience of young black males.

good are competitive. There are no state-contingent markets to insure against income risk, but workers can self-insure by borrowing and saving into a risk-free asset subject to a borrowing constraint. We denote by $\underline{a}$ the economy-wide, exogenous borrowing limit and by $\underline{a}_j^{nat}$ the natural borrowing limit.[15] There are perfect annuity markets to insure against mortality risk.

We normalize the price of the final good to one and denote by w^e the post-tax price of an effective unit of labor of type e and by r the post-tax riskless interest rate.

2.3.1 Work and crime

Crime and apprehension technologies: For a victim, a theft involves losing a fraction κ of labor income. It is modeled as a multiplicative shock $v \in \{0, \kappa\}$ to labor income with $\pi_v = Pr\{v = \kappa\}$ the probability of being a victim. For simplicity we assume criminals cannot target their victims and that each theft yields a fraction κ of the average labor earnings.

A criminal committing τ_j thefts is apprehended and sent to prison with probability $\tau_j \pi_p$. Convicted criminals receive no labor income and cannot be robbed, but keep their assets and the proceeds from their last crime. While in jail, they cannot access their assets. To streamline notation, we assume – just in the exposition – that the length of the prison term is one period.[16]

Earnings: The legal earnings for an individual of age j, education e and ability θ are given by $w^e h_j(\theta, e, \varepsilon_j^e)$ where

$$\log h_j(\theta, e, \varepsilon_j^e) = \gamma^e \theta + \zeta_j^e + \varepsilon_j^e, \tag{5}$$

where γ^e is an education-specific ability gradient, ζ_j^e an education-specific deterministic age component and ε_j^e a stochastic component following the process

$$\varepsilon_j^e = \rho^e \varepsilon_{j-1}^e + \eta_j^e, \qquad \eta_j^e \overset{iid}{\sim} N(0, \sigma^e). \tag{6}$$

[15]The exogenous borrowing limit $\underline{a}$ is tighter than the natural borrowing limit $\underline{a}_j^{nat}$ until close to the end of life. For this reason, in the exposition, we apply the general borrowing limit $\max\{\underline{a}, \underline{a}_j^{nat}\}$ only to retired workers.

[16]We do not impose the restriction in the quantitative analysis.

Individuals can engage in criminal activity while working. An individual's dynamic budget identity satisfies

$$a_{j+1} = a_j(1+r)\lambda_j^{-1} + \tau_j \kappa \overline{wh} + (1-i^p)[(1-v)\, w^e h_j \left(\theta, e, \varepsilon_j^e\right) - c_j], \tag{7}$$

where a_j is the stock of financial wealth, r the (post-tax) risk-free interest and i^p is a random variable which takes value one if $\tau_j > 0$ and the individual is convicted, and zero otherwise. The term λ_j^{-1} is the adjustment for the actuarially-fair annuity premium.

The first two addenda on the right hand side of equation (7) are common to all individuals, whether apprehended or not in the current period. The term $\tau_j \kappa \overline{wh}$ is the illegal return from committing τ_j crimes, with $\overline{wh}$ being average labor earnings. In addition, individuals who are not in prison in the current period – $i^p = 0$ – earn labor income $w^e h_j \left(\theta, e, \varepsilon_j^e\right)$, can be robbed of a share v of it and choose consumption c_j.

2.4 Government and education cost

Government: The government administers a pay-as-you-go pension system, the criminal justice system, spends on wasteful public expenditure and transfers, and collects taxes. Namely, it pays a pension benefit pen to each pensioner and bears a total cost m for each convicted criminal. It also pays a subsidy[17] sub^e for each period spent studying towards a degree e. Both pension benefits and student subsidies are tax-exempt while labor and capital income are taxed at the proportional rates t_l and t_k respectively.

The government balances the budget at all times. In the model benchmark, once the transfers and the criminal justice systems have been financed, any excess tax revenue is spent on non-valued public expenditure G.

Education costs: The out-of-pocket cost of studying toward degree e equals the student fee f^e minus the government subsidy sub^e for each year attended.

[17]For expositional convenience, we do not allow for means-tested subsidies in the description of the model, though we do allow for them in the numerical experiment in Section **??**.

2.5 The individual problem

We write the individual problem in recursive form. Given that the innate ability θ and the utility cost χ of engaging in crime are permanent individual characteristics, we subsume them in the value functions in what follows.

Education stage: Let $V_j^s(e, a_j)$ and $V_j^n(e, a_j)$ denote respectively the value of being or not in education for an individual of age j, (omitted) type (θ, χ), completed education e and financial wealth a_j.

The value function of a newborn with initial wealth a_0 satisfies

$$V^0(L, a_0) = \max\{V_0^s(L, a_0), V_0^n(L, a_0)\}, \tag{8}$$

as the agent chooses optimally between attending high school and entering the labor market as a high school dropout.

If $j \neq j_H - 1$, students with degree e studying towards degree e' solve

$$V_j^s(e, a_j) = \max_{c_j, a_{j+1}} u(c_j) + \psi^{e'}(\theta) + \xi \lambda_j V_{j+1}^s(e, a_{j+1}) \tag{9}$$

$$\text{s.t. } a_{j+1} = a_j(1+r)\lambda_j^{-1} + f^{e'} - sub^{e'} - c_j, \quad a_{j+1} \geq \underline{a}. \tag{10}$$

These students do not face a study-work choice at age $j+1$ and their continuation value is V_{j+1}^s. Among them, college students aged $j_C - 1$ will enter the labor market the following period and their continuation value satisfies the condition $V_{j_C}^s(C, a_{j_C}) = V_{j_C}^n(C, a_{j_C})$.

On the other hand, students in the last year of high school ($j = j_H - 1$) solve

$$V_j^s(e, a_j) = \max_{c_j, a_{j+1}} u(c_j) + \psi_{e'}(\theta) + \xi \lambda_j \max\{V_{j+1}^s(e', a_{j+1}), V_{j+1}^n(e', a_{j+1})\} \tag{11}$$

$$\text{s.t.} \quad (10).$$

Their continuation value is the result of the optimal choice, next period, between continuing education or entering the labor market as a high school graduate.

Work stage (no-bequests): Let the superscript p (for prison) index a convicted criminal and let $\mathbb{E}_{\varepsilon_j^e}$ and $\mathbb{E}_v$ denote the expectation operators with respect to the probability distribution of the labor efficiency shock ε_j^e and the victimization shock v. The value function of a worker aged $j \neq j_b$, before observing her current labor efficiency shock, is

$$V_j^n(e, a_j) = \mathbb{E}_{\varepsilon_j^e}\left[\max_{\tau_j} \ \chi \mathbb{I}_{\tau_j} + \tau_j \pi_p V_j^p(e, a_j) + (1 - \tau_j \pi_p)\mathbb{E}_v[\max_{c_j, a_{j+1}} u(c_j) + \xi \lambda_j V_{j+1}^n(e, a_{j+1})]\right] \tag{12}$$

$$\text{s.t.} \qquad (7), \qquad a_{j+1} \geq \underline{a}.$$

Upon observing her efficiency shock ε_j^e, an individual chooses how many crimes to commit. If $\tau_j > 0$, she bears a utility cost χ and is apprehended with probability $\tau_j \pi_p$. Any individual not in jail is subject to the random shock v associated with being robbed and chooses consumption after observing the shock realization. The value function of an individual in jail is

$$V_j^p(e, a_j) = u(\bar{c}) + \xi \lambda_j V_{j+1}^n(e, a_{j+1}). \tag{13}$$

Work stage (bequest age): An individual out of jail at the bequest age $j = j_b$ solves

$$\begin{aligned} V_j^n(e, a_j) = \mathbb{E}_{\varepsilon_j^e}\Bigg[\max_{\tau_j} \ & \chi \mathbb{I}_{\tau_j} + \tau_j \pi_p V_j^p(e, a_j) + \\ & (1 - \tau_j \pi_p)\mathbb{E}_v[\max_{c_j, a_{j+1}, b} u(c_j) + v(b) + \xi \lambda_j V_{j+1}^n(e, a_{j+1})]\Bigg] \\ \text{s.t.} \ \ a_{j+1} &= a_j(1+r)\lambda_j^{-1} + \tau_j \kappa \bar{w} + (1 - i^p)[(1 - v)\, w_j\left(\theta, e, \varepsilon_j^e\right) - c_j - b], \\ a_{j+1} &\geq \underline{a}, \quad b \geq 0. \end{aligned} \tag{14}$$

The constraint $b \geq 0$ rules out the possibility of parents extracting resources from their children.

An individual in jail at age j_b leaves no bequests and has value function

$$V_j^p(e, a_j) = u(\bar{c}) + v(0) + \xi \lambda_j V_{j+1}^n(e, a_{j+1}). \tag{15}$$

Retirement stage: From age j_r until death, individuals solve the problem

$$V_j^n(e, a_j) = \max_{c_j, a_{j+1}} u(c_j) + \xi \lambda_j V_{j+1}^n(e, a_{j+1}), \tag{16}$$
$$\text{s.t. } a_{j+1} = a_j(1+r) + pen - c_j, \quad a_{j+1} \geq \max\{\underline{a}, \underline{a}_j^{nat}\}.$$

3 A simple analytical framework

Before turning to the numerical analysis we use a stripped-down version of the model, which can be solved analytically, to highlight the basic economic forces at work.[18]

There are only two periods and the population has measure one. Agents differ in their ability θ, with θ distributed uniformly on $[0, 1]$. There are two education levels $\{L, H\}$, low and high. In the first period agents choose whether to enjoy leisure, in which case their education level in the second period is L, or go to school to acquire education level H. The utility of leisure is normalized to zero, while the cost of studying is $d - \log(\theta)$.[19]

In period two, agents inelastically supply one unit of labor for wage $w^e\theta$ and draw an additive taste shock for crime, u, uniformly distributed in $[-a, a]$. After observing the shock, agents optimally decide whether to engage or not in crime. With probability π, criminals are apprehended, lose their labor income and go to jail, where they consume $\bar{c}$. There are no capital markets, the intertemporal discount factor is one and the felicity from consumption in the second period is logarithmic.

The second period problem for an agent of education e and ability θ is

$$U_2^e(\theta) = \max\left\{\log\left(w^e\theta\right), u + (1-\pi)\log\left(w^e\theta\right) + \pi \log \bar{c}\right\}. \tag{17}$$

If the agent does not engage in crime, she consumes her labor income $w^e\theta$ with probability one. If instead she does engage in crime, she gets the utility from crime u plus the expected utility from consuming her labor income (if she is not apprehended) or the exogenous amount $\bar{c}$ (if she is apprehended).

[18]Here we only discuss the main results. See Appendix A.1 for the details of the derivations.

[19]The term d captures the total, ability-independent, part of the cost of studying, including tuition fees.

An agent engages in crime if and only her draw of u is above the reservation value

$$u_r^e(\theta) = \pi(\log(w^e\theta) - \log \bar{c}), \tag{18}$$

which implies that the probability of engaging in crime is decreasing in the probability of apprehension and, through the opportunity cost $\log(w^e\theta) - \log \bar{c}$, in the ability and education levels.

The first-period expected utility for a student of type θ is

$$U_1^H(\theta) = sub - d + \log(\theta) + \mathbb{E}U_2^H(\theta),$$

where sub is an education subsidy (in units of utility), $d - \log(\theta)$ the first-period disutility of studying and $\mathbb{E}$ the expectation operator. The first-period expected utility of an agent who does not study is $U_1^L(\theta) = \mathbb{E}U_2^L(\theta)$.

If $\frac{\pi^2}{4a} \simeq 0$ (see Appendix), the marginal ability θ^*, for which an agent is indifferent between studying or not, satisfies

$$\log \theta^* \simeq d - sub - \left(1 - \frac{\pi}{2}\right) \log \frac{w^H}{w^L}. \tag{19}$$

Note that θ^* is also the proportion of non-educated workers, given the wage premium.

The production technology is $Y = H = [sH_L^\varrho + (1-s)H_H^\varrho]^{\frac{1}{\varrho}}$, with H_L, H_H the stocks of the two worker types and $\rho \leq 1$. It follows that the skill premium can be written as

$$\log \frac{w^H}{w^L} = \alpha + (1-\varrho) \log \left(\frac{H_L}{H_H}\right) = \alpha + (1-\varrho) \log \left(\frac{\theta^{*2}}{1-\theta^{*2}}\right), \tag{20}$$

where $\alpha = \log[(1-s)/s]$[20] and the second equality follows from $H_L = \int_0^{\theta^*} \theta d\theta = \frac{\theta^{*2}}{2}$ and $H_H = \int_{\theta^*}^1 \theta d\theta = \frac{1-\theta^{*2}}{2}$. Substituting equation (20) into (19) gives the equilibrium share of unskilled workers

$$\log \theta^* \simeq d - sub - \left(1 - \frac{\pi}{2}\right) \left[\alpha + (1-\varrho) \log \left(\frac{\theta^{*2}}{1-\theta^{*2}}\right)\right],$$

[20]Note that if $s < 1/2$ (as it is surely the case empirically) we get that $\alpha > 0$.

which can be differentiated to obtain the general equilibrium response of the share θ^* to the high school subsidy

$$\frac{d\theta^*}{dsub} = -\theta^* \left[\frac{1-\theta^{*2}}{1-\theta^{*2}+(2-\pi)(1-\varrho)} \right]. \tag{21}$$

The general equilibrium change in the share of low educated workers is the product of two terms. The first is the partial equilibrium response $-\theta^*$ (cfr. equation (19)), while the second term captures the dampening general equilibrium effect due to the fact that an increase in θ^* increases the skill premium and, therefore, the incentive to acquire education. This second effect is decreasing in ϱ and absent in the case in which $\varrho = 1$, when the skill premium is independent of the relative supplies of skills. It follows that the absolute value of $d\theta^*/dsub$ increases as ϱ approaches 1.

In equilibrium, the probability that a worker of type (e, θ) is a criminal is

$$\Pr\{u \geq u_r^e(\theta)\} = \frac{1}{2a} \left[\bar{u} - \pi \left(\log w^e + \log \theta - \log \bar{c} \right) \right].$$

By integrating over ability and the two education levels one obtains the following expression for the total measure of criminals

$$C = \frac{1}{2a} \left[a - \pi \int_0^1 \log \theta d\theta - \log \bar{c} + \pi \theta^* \log \left(\frac{w^H}{w^L} \right) - \pi \log w^H \right]. \tag{22}$$

Note that an increase in the share of uneducated workers increases the number of criminals through three effects. The first is a (partial equilibrium) composition effect: an increase in θ^* increases the number of criminals, given a positive skill premium, because unskilled workers have a lower opportunity cost of crime. The second, general equilibrium effect, reinforces the first one. By increasing the skill premium $\log(w^H/w^L)$ an increase in θ^* reduces the opportunity cost of crime for non-educated workers, for given w^H. Finally, to the extent that an increase in θ^* increases w^H, it also decreases the crime rate among educated workers, partly counteracting the other two effects. The two general equilibrium effects decline in size as ϱ increases, and disappear when $\varrho = 1$. Therefore, for a given initial θ^*, the difference between the partial and the general equilibrium derivative of C with respect to θ^* is a decreasing function of ϱ.

The general equilibrium change in the number of criminals due to an education subsidy is the product of $\partial C/\partial\theta^*$ and the derivative in equation (21). This product is 'a priori' ambiguous in sign. However, in the numerical analysis we consistently find that, for all parameterizations, the change in w^H is small relative to the change in the overall skill premium: this implies that $\partial C/\partial\theta^*$ is positive and the equilibrium effect of a subsidy on crime is negative. Note also that, as ϱ changes, the absolute values of $\partial\theta^*/\partial sub$ and $\partial C/\partial\theta^*$ move in opposite directions. Therefore their product might not change dramatically with skill substitutability. In other words, the equilibrium effect of a subsidy on crime might not be extremely different for different values of the elasticity of substitution in skill types.

Nonetheless, this simple model reveals that the crime response to a subsidy policy should change with ϱ and, through the *level* of the skill premium, with the human capital shares in production.

4 Parameterization

We now turn to the description of the model calibration. We begin with the parameters set outside the model and then discuss those whose calibration requires solving for equilibrium. For the latter set of parameters, calibration is obtained through minimization of the sum of squared deviations of simulated and data moments. The first set of parameters is reported in Table 1 while the second set is listed in Table 2. The table also reports the set of targeted moments. The last line in Table 2 lists a moment but no parameter, as the number of target moments exceeds the number of parameters by one.

We calibrate the model to the early Eighties. All income flows in the code are expressed as a share of average labor earnings in 1980.[21] We report them in 1980 dollars in what follows.

Demographics. Each period represents one year. The real world counterpart of the first year of age in the model ($j = 0$) is age 16 and the last model year ($j = \bar{j}$) corresponds to age 95. The inter-vivos bequest takes place at age 45 and retirement at age 65. The

[21] This amounted to 12,400 dollars.

survival probabilities λ_j equal one until the bequest age and are taken from NCHS (1997) for subsequent ages.

Production technology. The capital depreciation rate δ and the share of capital income ϕ are set respectively to 0.065 and 0.35 (see Cooley, 1995). The parameters ϱ, s_L and s_H of the human capital aggregator in equation (4) are based on estimates by Abbott, Gallipoli, Meghir, and Violante (2012) using CPS and PSID data. For ϱ, we use their favorite estimate of $\varrho = 0.68$, corresponding to an elasticity of substitution of 3.1.[22] The share parameters are calibrated using an average of the values estimated for the period 1979-1981 and are report in See Table 3.

Government. The government levies proportional taxes on capital and labor. Following Domeij and Heathcote (2003), we set $t_l = 0.27$ and $t_k = 0.4$. For simplicity, the pension is assumed to be a constant lump sum for all agents, regardless of their education and previous earnings. The pension *pen* is set to 1,980 dollars which corresponds to a pension replacement rate of 16 per cent of average labor earnings as in Heathcote, Storesletten, and Violante (2010).

The expenditure per convict in the model is set to $8,300. This corresponds to the average (across jails and state prisons) cost per prisoner in 1980.[23]

Education subsidies sub^e are set to zero in the benchmark.

Education duration and costs. Consistently with our assumption that model age $j = 0$ corresponds to age 16 in the data, obtaining a high school degree requires two years of attendance – $j_H = 2$. Obtaining a college degree requires four more years of attendance, which implies $j_C = 6$.

The direct cost of college education is chosen to match the value of the average tuition costs net of the average grant for the academic year 1980-81. The figure for the average

[22]Estimates of the degree of substitutability between high school dropouts and other skill groups are also provided by Goldin and Katz (2010), using U.S. Census data for almost all the 20th century. They place its value in a range between 2 and 5. Katz and Murphy (1992) and Heckman, Lochner, and Taber (1998) estimate the elasticity of substitution between college graduates and non-graduates to be respectively 1.41 and 1.44 while Card and Lemieux (2001) find that the elasticity of substitution between different age groups is large but finite (around 5) while the elasticity of substitution between college and high school workers is about 2.5.

[23]The figures for the average cost per jail and state prisoner are respectively from "Correctional Populations in the United States" (NCJ-156241), 1993 and "Federal and State Prisons. Inmate Populations, Costs and Projection Models" (B-272244), General Accounting Office, 1996.

tuition cost is \$1,679, the average "Tuition and required fee" over all 4-year institution from the National Center for Education Statistics. The average grant size for the 60 per cent of the students who received it is \$795, according to Lewis (1989). This implies an unconditional average grant of \$478 and an average yearly net cost of \$1,200. As for the yearly cost of attending high school, we set it to be just \$124 (or 1% of average labor earnings), in order to account for expenses incurred for study material and other costs. There does not seem to be much information on such costs, therefore we also experiment by considering alternative values in Section 6.

Legal earnings. It follows from equation (5) that the model implies the following specification for the legal earnings $W^e_{ijt} = w^e h^e_j(\theta_i, e, \varepsilon^e_{ijt})$ of an individual i, with education e, at age j and time t

$$\log W^e_{ijt} = \log w^e + \gamma^e \theta_i + \zeta^e(j_{it}) + \varepsilon^e_{ijt} \tag{23}$$

The values of all the parameters in equation (23) are taken from the estimates in Abbott, Gallipoli, Meghir, and Violante (2012). The estimates for the ability gradient[24] are based on NLSY79 data and reported in Table 4. The estimates for the age component $\zeta^e(j_{it})$ are based on PSID data, which provide a longer working life span, and are presented in Table 5.

The estimates for the parameters ρ^e and σ^e of the stochastic component

$$\varepsilon^e_{ijt} = \rho^e \varepsilon^e_{i,j-1,t-1} + \eta^e_{ijt}, \qquad \eta^e_{ijt} \overset{iid}{\sim} N(0, \sigma^e) \tag{24}$$

are reported in Table 6.

Preferences. The felicity function over consumption is CRRA. We set the coefficient of relative risk aversion to 1.5, in the middle of the range of available estimates (see, for example, the survey by Attanasio, 1999).

We set the discount factor ξ to match a ratio of average net wealth to average income of 2.7. estimated from the 1983 Survey of Consumer Finances as reported in Wolff (2000)[25]

[24]The measure of ability, both in the model and in the data, is the deviation of the log of the AFQT89 test score from its mean.

[25]It is well known that the PSID and NLSY, from which most of our other estimates are obtained, signif-

The implied value for the discount factor is 0.967.

The warm-glow utility from bequests has the same functional form

$$v(b) = \nu_1 \left(\nu_2 b\right)^{\nu_3} \tag{25}$$

used by De Nardi (2004).

To calibrate $\{\nu_1, \nu_2, \nu_3\}$ we target three moments of the distribution of inter-vivos transfers in the data. Data on inter-vivos transfers to young individuals are not available in the NLSY79, but the NLSY97 contains information on family transfers received by young individuals. For this reason, we use statistics from the NLSY97 and convert them into 1980 dollars. Since in the model inter-vivos transfers are one-off bequests which are mainly relevant for the education choice, we restrict attention to transfers received between 16 and 22 years of age in the NLSY97. The targets are computed using the statistics reported in Abbott, Gallipoli, Meghir, and Violante (2012).[26] The three moments we target are the total (over the age range 16-22) average inter-vivos transfer, its coefficient of variation and the share of individuals who receive no transfer. They are respectively 15,200 dollars, 0.75 and 9 per cent.

Borrowing Limit. The exogenous borrowing limit $\bar{a}$ is calibrated to match the share of workers (all agents excluding students) with zero or negative wealth. Wolff (2000) provides an estimate of 15% for this share in 1983, which implies a negative borrowing limit of about 3,960 dollars.

Disutility of schooling. The terms $\psi^H(\theta)$ and $\psi^C(\theta)$ are calibrated to match the fractions of high-school and college graduates in the five ability bins in which we have partitioned the range of the AFQT89 scores from the NLSY79. Since the NLSY79 provides also information on educational attainment, it allows to estimate the joint distribution of ability and education. However, the aggregation of the education shares based on the NLSY79 is not consistent with the aggregate education distribution of workers in the

icantly undersample rich households relative to the SCF. For this reason, the value of the wealth/income ratio we use is calculated across all households excluding the top 5 per cent. Kaplan and Violante (2010) make a similar argument.

[26]The transfers also include imputed rent for students' living with their parents. More details are available upon request.

U.S. economy, as measured by the CPS for 1980.[27] For aggregate consistency we adjust the NLSY79 rates so that they aggregate to the average CPS education rates observed between 1977 and 1983.[28] The education shares by ability and their aggregate value are reported in Table 7.

Crime parameters. Our definition of property crime is the same as in the data. It is any crime which qualifies as burglary, larceny or motor-vehicle theft. Our measures of income from crime, probability of incarceration and length of sentence are averages of the corresponding measures in each of these categories with weights equal to the relative frequency of each of them.[29]

We calibrate κ, the share of income lost if victimized, to match the average loss for a victim of one property crime in the data. From Table 3.76 in Pastore and Maguire (1982), the average property crime loss was \$728 in 1980. The resulting value for κ is 5.87 per cent.

The probability of being convicted for committing one crime is the product of the probability that a crime is cleared by arrest and the suspect indicted (clearance rate), the probability of facing trial conditional on being indicted and the probability of being sentenced to prison conditional on being put on trial. These probabilities were respectively 16.8, 80 and 44 per cent in 1980[30] and imply $\pi_p = 5.7$ per cent.

The sentence length is set to 19 months, which is the average time served in 1983, the closest survey year, by prisoners convicted for property crime offenses released from state prisons.[31] Following Cozzi (2006), we model the fractional prison term as a lottery between a prison term of one and two years with respective probabilities that imply an expected prison sentence of 19 months.[32] The reason for modeling fractional prison terms

[27]The reason for this discrepancy is that the NLSY79 refers only to one cohort of the U.S. population, whereas the CPS gives a snapshot of the education distribution of workers at all ages.

[28]The aggregate education distribution for workers between 1977 and 1983 was close to the average for the period 1967-2001, which is the sample period used in Abbott, Gallipoli, Meghir, and Violante (2012) to estimate the wage equations and the production technology.

[29]The weights change with the variable under consideration; e.g. they are the relative frequency of each category of property crime in the case of income from crime, but the relative number of prisoners charged with each category of crime in the case of sentence length.

[30]Respectively from tables 4.19, 5.19 and 5.20 of Pastore and Maguire (1981).

[31]From Table 6.31 in Pastore and Maguire (1986).

[32]This implies that an individual committing crime in the last year of her working life spends the first retirement year in prison with positive probability. We assume that her pension is paid into her bank account, though he cannot access it until released.

this way is that reducing the period length to an appropriately small unit (e.g. one or two months) would have increased computational costs significantly. On the other hand, maintaining a yearly time unit while allowing individuals to work and be victimized for a fraction of a year after coming out of jail is extremely cumbersome.

We set the exogenous consumption level in prison $\bar{c}$ to match a property victimization rate of 5.6 per cent in 1980.[33] The resulting value is $\bar{c} = 3,900$ dollars.

Individual fixed effects. In the model ability θ is a set of innate and permanent characteristics which affect earnings and educations choices. To this purpose we proxy ability by the deviation of the log of the AFQT89 test score (as reported in the NLSY79) from its mean.[34] For computational simplicity, the distribution is approximated by grouping all agents in 5 bins (quintiles) containing equal proportions of the total population. The range of each bin is different, as its extremes correspond to successive quintiles of the empirical ability distribution. Ability is assumed to be uniformly distributed within each bin.

The counterpart of the crime fixed effect χ in our model is unobservable in the data and we have no prior on its distribution. For this reason we assume that it follows a Beta distribution – one of the most flexible parametric distributions – with support $(\underline{\chi}, 1-\underline{\chi})$ and exponents $\{\alpha, \beta\}$.[35]

We allow for an arbitrary correlation $\rho_{\theta\chi}$ between the ability and crime fixed effects. We do so through a normal copula.[36]

To calibrate the four parameters $\{\underline{\chi}, \alpha, \beta, \rho_{\theta\chi}\}$ we target the following five moments. The first four are obtained by running, in the spirit of indirect inference, the same regression on both simulated data and the 1980 wave of the NLSY79 for individuals aged 18 to 23. These data contain the same information on AFQT89 that we have used to calibrate our model and, more importantly they contain specific information on self-reported participation in property crime.[37] The measure of property crime we use is whether or

[33] The figure is from Table 3.53 in Pastore and Maguire (1982).

[34] This data set has the advantage of providing a measure of cognitive skills for a representative sample of an entire cohort in the U.S. population, as well as direct wage measures over time. We can therefore link measured 'ability' and wages within different education groups.

[35] Depending on the value of its exponents the Beta distribution encompasses the uniform, the truncated normal, the exponential and allows for asymmetry and fat tails.

[36] Nelsen (2006) is a standard reference on the topic.

[37] The same dataset has been used by Lochner (2004) and Lochner and Moretti (2004) to study the

not the respondent reported having engaged in shoplifting, or stealing something worth $50 or more, from someone/somewhere other than a store. The counterpart in the model data is whether an individual commits a positive number of crimes in a given period.

The regression dependent variable is a dummy equal to 1 if the individual is engaged in crime in the current period and zero otherwise. The independent variables are a constant, age, the AFQT89 percentile and an education dummy equal to 1 if the individual has at least a high school degree and zero otherwise.

The fifth targeted moment is the share of state prison inmates who do not have a high school degree. Its value in 1979 was 52.7.[38]

The four model parameters determine the relative contribution of the economic and non-economic crime motive in the model. We note that the calibrated value of $\rho_{\theta\chi}$ turns out to be 0.12, meaning that relatively more able people have on average a very slightly higher propensity to engage in crime. All other parameters are reported in Table 2.

Discussion of identification: While in theory all calibrated parameters are *jointly* identified by the chosen moments, in practice some parameters affect only a subset of the selected moments, so that identification is effectively block recursive. Here we briefly discuss the relation between moments and parameters.

Education moments. A first set of ten parameters captures the non-pecuniary costs of acquiring education $\{\psi^H(\theta), \psi^C(\theta)\}$. These parameters are chosen to match the benchmark education distribution within the five ability bins, both for high school and for college.

Wealth moments. A second set of parameters, consisting of the subjective discount factor ξ, the three parameters $\{\nu_1, \nu_2, \nu_3\}$, indexing the warm-glow utility from bequest $v(b) = \nu_1(\nu_2 b)^{\nu_3}$, and the borrowing limit $\underline{a}$, is identified by moments of the wealth distribution. Of these, $\{\nu_1, \nu_2, \nu_3\}$ mostly affects the three targeted moments of the distribution of inter-vivos bequests. In particular, ν_2 is pinned down by the share of individuals receiving zero bequests and, to a lesser extent, by the variance of bequests. The parameter ν_3 changes both the variance and the average bequest, while ν_1 mostly influences the av-

effect of education on crime. The dataset we use contains individuals aged 18-23 and has been kindly provided by Lance Lochner. A large subset of the data is available on Enrico Moretti's webpage at http://www.econ.berkeley.edu/ moretti/data.html

[38]The figure is from "Profile of State Prison Inmates" (NCJ-58257), 1979.

erage bequest. The discount factor ξ affects both the aggregate wealth-income ratio and the average bequest. Finally, given the values of the other four parameters, the borrowing limit $\underline{a}$ is used to target the share of agents with non-positive wealth.

Crime moments. While the parameters $\{\xi, \nu_1, \nu_2, \nu_3, \underline{a}\}$ affect the target crime moments by changing the wealth distribution, the reverse is not true. The remaining set of parameters, $\{\underline{\chi}, \alpha, \beta, \rho_{\theta\chi}, \kappa, \bar{c}\}$, affect mainly the seven crime moments and are (over-)identified by them for a given value of the other parameters.

In particular, $\underline{\chi}$ determines the location of the distribution of the crime fixed effect. An increase in this parameter results in a larger share of 'non-economic' versus 'economic' crime. Since most of the economic component of crime is concentrated among agents with low ability and education this parameter is identified by the share of criminals who are high school dropouts.

The parameter $\rho_{\theta\chi}$ determines the correlation between ability and the crime fixed effect and, through selection on ability into education, it also helps determine the crime rate among high school graduates. For this reason, it is identified by the coefficient of education in the crime regression used for the calibration.

Given that the two parameters above determine the location of crime fixed effects and their correlation with ability, we are left with the parameters (α, β) which determine the shape of the distribution of crime fixed effects. These shape parameters shift the probability mass to, respectively, the left and right tail of the distribution. By doing so they increase the proportion of inframarginal non-criminals (left tail) and criminals (right tail) and they affect the proportion of marginal criminals whose choices are driven by market returns which, for given education, depend on age and ability. For this reason, they are identified by the ability and age coefficients in the calibration regression.

5 Numerical Analysis

This section discusses the benchmark equilibrium and presents the results of our policy experiments. Section 5.1 discusses the implications of the model for a range of non-targeted data moments. The remaining sections describe policy experiments. In all the experiments, the proportional tax rate on labor income adjusts to balance the government budget. An extensive sensitivity analysis is reported in Section 6 while Appendix A.3 contains a discussion of the numerical solution method employed.

5.1 Model performance

The model does a good job matching the targeted moments (see Table 2). It is worth noticing that the share of high school dropouts among prisoners, as well as the constant and education coefficients in the crime participation regression, are lower than the targets. For this reason, our analysis is likely to provide a lower bound of the true impact of schooling on the victimization rate.

The model has also implications for a range of untargeted moments, which can be used to assess its performance. In particular, we look at the model's prediction for the age profile of arrest rates, the impact of education on incarceration rates and the response of high school and college enrollment to subsidies. We also evaluate the joint response of education and crime rates to various changes occurring between 1980 and 2000; we consider changes in skill-bias in production, the severity of punishment, labor market risk, tuition costs and demographics. Table 8 reports all these validation results, comparing moments from the data and the model.

Age profile of arrest rates. Figures 1 and 2 report the age profiles of arrest rates (for property crimes) between age 18 and 60 in both the model and the US data[39]. The data are available for each individual age until age 24 and for five-year age brackets (e.g. 25-29) from age 25 onwards. Arrest rates decline with age in both data and model. The model provides quite a good match until middle-age, but overstates arrest rates after age 45.

Relationship between education and crime. The model predicts that education should reduce crime participation by increasing the opportunity cost of crime. There is substantial empirical evidence that increases in market returns reduce crime participation. In particular, Lochner and Moretti (2004) provide evidence on the causal effect of education on various measures of criminal activity. They estimate the impact of high

[39] From Table 4.4, Sourcebook of Criminal Justice Statistics (1980).

school graduation on the probability of being incarcerated for men aged 20 to 60 using Census data for the years 1960, 1970 and 1980 (see their Table 9). Their preferred estimates are obtained by instrumenting high school graduation through changes in compulsory schooling laws.[40] Since our model is calibrated to the year 1980, we use their same approach (and their Census data) to estimate the effect of high school graduation for the year 1980. Our model does not disaggregate crime by ethnicity, so we estimate the same regression using their pooled sample and obtain an estimated coefficient of -0.47 (standard error 0.45).[41] Finally, running the same regression on model-generated data and using the same set of common controls,[42] plus the crime fixed effect to directly control for the unobserved crime heterogeneity, we obtain an estimate of -0.37.

Lochner and Moretti (2004) also estimate the impact of high school graduation on self-reported participation rate in property crime using the 1980 wave of the NLSY79 for individuals aged 18 to 23 (see Table 12 in their paper). Estimating their regression on their pooled sample produces a coefficient of -5.63 (standard error of 2.25). Estimating the same regression using simulated data, with the same set of common controls plus the crime fixed effect[43], we obtain an estimate of -4.07. We also experiment with controls for earnings in the simulated data regression: this does not affect the estimated coefficient because age, ability and education are enough to capture the opportunity cost of crime in the model.

Enrollment response. The model response of high school and college enrollment to conditional cash transfers appears to be consistent with available reduced form studies. Since, in general, such studies concern small-scale quasi-experimental interventions, we compare their results to a partial equilibrium numerical experiment under the assumption that the policy change is unexpected, while keeping the initial wealth distribution unchanged.

[40] Lochner and Moretti notice that "...OLS estimates may reflect the effects of unobserved individual characteristics that influence the probability of committing crime and dropping out of school. For example, individuals with a ... taste for crime...are likely to commit more crime and attend less schooling."

[41] When we estimate the regression, like Lochner and Moretti, for white men, the estimated coefficient is -0.39 (standard error 0.36).

[42] Namely, age, ability and the AFQT89 percentile.

[43] This regression is similar to the one used to calibrate the model, but in addition controls for the crime fixed effect to mimic the extensive set of proxies used by Lochner and Moretti to control for unobserved heterogeneity in criminal behavior.

Dearden, Emmerson, and Meghir (2009) measure the effect of subsidising attendance to (post-compulsory) high school: they use data from a UK pilot study (the EMA) which offered means-tested conditional cash transfers for 16-to-18 year olds to stay in full-time education. They find that such transfers increased the share progressing to two additional years of education by around 6.7 percentage points (with a standard error of 1.7). The average transfer was roughly 20 per cent of the median post-tax earnings of high school dropouts at age 16.[44] A similar subsidy in the model increases high school attendance by a remarkably similar 6.6 percentage points.

We also compare simulated college enrollment responses to a change in tuition costs to those estimated in the data. Kane (2003) reports a range of estimates for the college enrollment response to a $ 1,000 (in 2001 dollars) change in college tuition costs[45]. Such estimates range between 3 and 9 percentage points (in absolute value), with the majority of them in the 3 to 6 range. The response in the model is an increase of 3.6 percentage points in response to a reduction in tuitions and a fall of 4.9 points in response to an increase, both within the range of existing estimates.

Changes in crime and enrollment (1980 vs 2000). Imrohoroglu, Merlo, and Rupert (2004) were the first to introduce a dynamic, stochastic, equilibrium model of crime and used it to account for the evolution of property crime over the period 1980-96. In a similar spirit, here we assess the model ability to account for the *joint* response of crime and enrollment to changes in fundamental parameters. Taking the benchmark calibration for 1980, we input (both simultaneously and one at a time) data for the human capital shares, the apprehension probability and the length of the prison term, the variance of income shocks, college tuitions and the age composition of the labor force for the year 2000. For each change of parameters we compute the steady-state equilibrium for the model economy and compare the crime and enrollment rates in the model with those in the data.

Between the years 1980 and 2000:

[44] The transfer ranged between 30% and 5% of median post-tax earnings of high school dropouts at age 16. 47% of eligible people, those with parental income below £13,000, were entitled to the maximum payment, while 31%, with parental income between £13,000 and £30,000, were entitled to a reduced payment.

[45] 1,000 dollars in 2001 correspond to $420 in 1980.

- the high school and college shares changed respectively from 0.41 to 0.39 and from 0.37 to 0.45;[46]
- the probability of conviction for a property crime increased from 0.057 to 0.077, while the prison term for property crimes increased from 19 to 25 months;[47]
- the variance of income shocks increased by 23% according to Heathcote, Storesletten, and Violante (2010);
- college tuitions increased in real terms from 9.7% of average labor earnings in 1980 to 23% of average earnings in 2000;
- the population age composition saw an increase in average age; using Census data, we approximate the demographic change by a 26 per cent reduction in the 16-25 age bracket and respectively a 9, 12 and 10 per cent increases in the 26-45, 46-65 and 66-95 age brackets.

The model performs very well both along the crime and the enrollment dimension. When all changes are introduced simultaneously, the model generates a victimization rate of 3.6 per cent, which happens to be the same as in the data, and enrollment rates for high school and college of respectively 59 and 27.5 per cent compared to 60 and 26 per cent in the data. Table 8 also reports the model responses for each individual change in fundamentals: we report only moments which change from their 1980 counterpart. Overall the main drivers of changes in the victimization rate have been the change in 'between-groups' income inequality due to changes in skill bias and the increase in punishment (apprehension rates and prison term).[48] *Coeteris paribus,* the change in human capital shares would have increased the victimization rate to 8% from its baseline of 5.6 per cent in 1980, while the increase in expected punishment would have reduced it to 2.1%. The increase in the variance of income shocks and the change in demographics have effects which are similar in magnitude, but opposite in sign.

[46] Estimates by Abbott, Gallipoli, Meghir, and Violante (2012) available from the authors

[47] The data for 2000 are from Pastore and Maguire (2001). As discussed in Section 4, the probability of conviction is the product of the clearance rate, the probability of facing trial conditional on being indicted and the probability of being sentenced to prison conditional on being put on trial. These probabilities were respectively 16.7, 92 and 53 per cent according to tables 4.19, 5.17 and 5.19. The average prison term is from Table 6.38.

[48] This result is in line with the findings of Imrohoroglu, Merlo, and Rupert (2004) for the period 1980-96.

The change in enrollment rates was mainly driven by technological change in human capital shares, which was partly offset by the dramatic increase in college tuitions. The model suggests that, had it not been for the dramatic increase in tuitions, the change in the human capital shares would have implied a college enrollment rate of 32% rather than the observed 26%.

5.2 Subsidizing High School Completion

The first experiment we carry out involves subsidizing high school completion. The dramatic increase in the U.S. prison population over the past twenty years has prompted a shift of interest, among both academics and policymakers, from tougher sentencing to other forms of intervention.[49]

In the experiments in this section we consider a yearly subsidy equal to 8.8 per cent of average labor earnings paid to all targeted students attending and completing high school. The size of the subsidy corresponds to the average transfer of a well-documented experiment, conducted by the Department of Labor and the Ford Foundation, which was aimed at increasing the likelihood that participants would complete high school.[50] While the program had a substantial learning, support and mentoring component, which the model cannot capture, the aim of this exercise is to evaluate the effect of a realistically-sized subsidy policy. We assume the cost of the policy is financed by adjusting the labor tax rate.

5.2.1 Unconditional subsidy: partial equilibrium

In this subsection, we investigate the effect of a subsidy paid to *all* individuals completing high school independently of their financial means. We conduct the experiment in partial equilibrium, keeping all prices at their level in the benchmark economy. Unlike the partial equilibrium enrollment experiments in Section 5.1, we do allow the distribution of inter-vivos transfers and the aggregate victimization rate to change and we study

[49] See, e.g., Donohue and Siegelman (2004).

[50] The experiment - known as the Quantum Opportunities Program - was carried out on a small scale in two waves. A first ("Pilot") run took place between 1989 and 1993. A later ("Demonstration") run took place between 1995 and 2001. The value of 8.8% is roughly the ratio between $ 2150, the average transfer received by participants, and average labor earnings in 1995, a central year in the program.

the steady-state equilibrium in which the distribution of physical and human capital is stationary. This provides the appropriate counterpart to evaluate the general equilibrium experiment in the following subsection, isolating the direct effect of the subsidy on the victimization rate from its indirect effect due to a change in prices.

Table 9 reports the change in the education shares relative to the benchmark. The aggregate share of high school dropouts falls by 9 per cent. The fall affects all ability bins, but is largest for the lowest two. The significantly lower fall in the top three ability bins is accounted for by their lower share of high school dropouts in the benchmark. In fact, such share drops to zero as a consequence of the subsidy.

Column (2) in Table 11 reports a number of statistics for the benchmark equilibrium. Output and aggregate consumption (the net flow of consumable resources) are normalized to 100. The welfare criterion we employ is the permanent consumption level that would give an individual the same ex-ante (before ability, wealth and the crime fixed-effect are realized), expected lifetime utility of a newborn into the stationary equilibrium.

Turning to column (3) in Table 11, the victimization rate falls from 5.62 per cent in the benchmark to 5.44.

To understand what drives the change in the victimization rate it is useful to observe the change in the (unconditional) arrest rates by education, defined as the number of arrests in an education group as a share of the total number of individuals (criminals and non-criminals) in the group. Since the apprehension probability is linear in the number of crime committed, the arrest rate for each education group is proportional to the (unconditional) average number of crimes in the education group. The number reported is the actual rate multiplied by 10,000. The arrest rate increases respectively by slightly more and slightly less than 10 per cent for high school dropouts and high school graduates relative to the benchmark[51] This reflects the worsened ability composition in both education groups, due to selection into high school along the ability dimension.[52] As discussed in Section 3, the change in the education composition, though, reduces the aggregate victimization rate, as high school graduates, for any given ability, have a higher

[51] The crime rate of college graduates in negligible in the model.

[52] Compared to the benchmark, only people in the lowest two ability bins compose the pool of high school dropouts, while relatively less able people now obtain a high school degree. This is confirmed by the dramatic fall – from 48 to 36 per cent – in the share of high school dropouts relative to the benchmark.

opportunity cost of crime – hence a lower crime rate – relative to high school dropouts due to the higher price of their human capital.

The increase in welfare relative to the benchmark, as measured by the ex ante consumption equivalent, is a large 3.4 per cent. By weakening the link between wealth at birth and selection into education, the subsidy provides insurance against the (ex-ante) uncertain ability and initial wealth draws. Since the ability shock is permanent and the initial wealth shock has a persistent effect through the education choice, the welfare benefits are large as they cumulate over the whole lifetime.

Finally, prison expenditure falls marginally but total expenditure on prison plus education subsidies increases from 0.27 to 0.51 per cent of aggregate consumption in the benchmark. Despite this the labor tax rate (not reported) falls by half a percentage point due to the dramatic increase in the aggregate stock of human capital.

5.2.2 Unconditional subsidy: general equilibrium

We now investigate the effect of the same subsidy as in the previous subsection but in general equilibrium.

Table 10 reports the changes in the education shares relative to the benchmark economy.

As discussed in Section 3, the general equilibrium change in the education composition is smaller than in partial equilibrium: the share of high school dropouts increases by only 1 percentage point compared to 9 percentage points in partial equilibrium. The aggregate change is accompanied by increased sorting across ability bins relative to partial equilibrium. The share of high school dropouts actually *increases* in the lowest two ability bins while it falls in the others. Only in the top two ability bins is the fall in the number of dropouts the same as in partial equilibrium, a clear indication that more able people tend to crowd out less able ones when prices adjust.

Turning to Column (3) in Table 11, the shares of high school dropouts among criminals falls substantially less, 3 rather than 8 percentage points. Despite this significantly smaller composition effect, though, the victimization rate falls substantially more, from 5.60 to 5.19 per cent, an 8 per cent reduction. Prison expenditure falls by a similar percentage.

To understand what drives the difference, note that the arrest rates for both high school dropouts and graduates *fall*, rather than increase as in partial equilibrium, relative to the benchmark. The reduction is particularly pronounced for high school dropouts. The fall in the high school premium has increased the opportunity cost of going to high school and resulted in an improved composition of the pool of high school dropouts relative to partial equilibrium. This raises their average opportunity cost of crime at given labor prices and reduces their crime rate. This effect is reinforced by the general equilibrium increase in the price of their labor. The ability composition of the pool of high school graduates has also improved relative to partial equilibrium because now only individuals with relatively higher ability switch.

The improved education sorting by ability implies an increase in both output and aggregate consumption of roughly 1 percentage point and a marginal fall in the labor tax rate to 26.8 per cent.

The welfare change is only marginally higher than in partial equilibrium, implying that the small fall in inequality (the skill premium) has only minor effects on expected utility. Finally, the change in the total (prison plus subsidy) expenditure is smaller due to the smaller subsidy take up.

We also experiment with a subsidy equal to twice the original one (column (4) in Table 11). The increase reduces the aggregate share of high school dropouts by an additional 0.5 percentage points.

Though the marginal benefit (both in terms of crime reduction and of increased efficiency and welfare) is decreasing, reflecting the progressive exhaustion of the benefits from improved sorting into education, it is still substantial.

The victimization rate falls to just below 5 per cent while aggregate consumption and welfare increases substantially, by roughly 2 and 6.5 per cent relative to the benchmark.

Note that while the total expenditure increases with the size of the subsidy up to a sizable 0.71 per cent of aggregate consumption in the benchmark, the labor tax rate is hardly affected as the increase in the tax base generates the necessary increase in revenue.

5.3 Increasing the prison sentence

We now turn to the allocation and welfare effects of an increase in the prison term and compare them to the effects of the education subsidy analysed in the previous two sections. A natural way to compare these two policies is to consider an increase in the prison sentence that achieves *the same reduction* in the victimization rate as the education subsidy.

Columns (5)-(6) are meant to be compared with columns (3)-(4). Prison terms of, respectively, 20 and 20.5 months achieve the same victimization rates as subsidies equal to, respectively, 8 and 17 per cent of average labor income. The policy has basically no effect on the education distribution, the criminal composition, prices and aggregate output and consumption. Despite the same fall in the victimization rate, the total prison expenditure falls by less than in the case of the high school subsidy. The increase in the sentence length implies a higher stock of inmates for the same victimization rate.

Crucially, welfare hardly changes, as crime entails a pure redistribution in the model.

6 Sensitivity analysis

The analytical section highlights how the effects of subsidizing education on crime depend on the relative price and degree of substitutability in production between different education types. Moreover, initial resources, access to credit and education costs may also contribute to shaping policy responses. In what follows we examine alternative parameterizations where we perturb the technology parameters and the initial wealth distribution, which we find to be the most important determinant of the size of the equilibrium effects. We also report results from parameterizations in which we change credit constraints, education costs and the correlation of unobserved characteristics.[53]

For the technology parameters we conduct our experiments both in PE and GE because our model predicts that they are the main determinants of the GE effect. We study the same policy considered is Section 5.2.2, a high school subsidy equal to 8.8 per cent of average earnings. For each calibration we report the aggregate victimization rate and welfare in Table 12.

[53] All the alternative benchmarks are recalibrated, as described in Section 4.

Skill substitutability. We experiment with elasticity values equal to 2 and 5, as opposed to 3.1 in the benchmark economy. These correspond to the lower and upper bounds of estimates in Chapter 8 of Goldin and Katz (2010) for the elasticity between high school dropouts and graduates.

In both parameterizations the drop in crime is larger in general equilibrium than in partial equilibrium. As predicted by the analytical example, when the elasticity of substitution is smaller the fall in the victimization rate in PE is smaller, and the crime reduction due to GE effects is larger.[54] The welfare gains are very similar to the ones in our main parameterizations.

Human capital shares in production. A change in the skill bias of technology (as captured by a fall in the share of high school dropouts in production) is associated to an increase in the education premium. We simulate the model using the labour shares for the year 2000[55] and confirm the conjecture based on the analytical example: for a larger skill premium the same education subsidy generates a larger PE drop in the victimization rate: 5.2% rather than 5.44 in the benchmark. When general equilibrium effects are accounted for, the victimization rate drops even more, down to 4.9%. As the model predicts, the difference between PE and GE effects, which is driven by the unchanged elasticity parameter rho, is similar to that in the benchmark parameterization.

Initial wealth distribution. One crucial source of heterogeneity and selection in the benchmark economy is the initial asset distribution. We experiment with changing the parameters η_1 and η_2 of the warm glow utility from bequests, which shape the wealth distribution at the start of life. Specifically, we change η_1 by the amount necessary to alter the average inter-vivos transfer by plus and minus 5,000$, relative to 15,200$ in our benchmark calibration. We change η_2 to alter the proportion of agents born with zero wealth, to respectively 4% and 14% as opposed to 9% in our benchmark calibration. As expected, increasing the average inter-vivos transfer and reducing the share of individuals born with zero wealth decreases the impact of the subsidy policy both on the victimization rate and on welfare, relative to the main calibration. The reverse is true for a reduction

[54] The victimization rate does not change in PE in the low-elasticity case. As in the analytical example, the very small high school wage premium implies a small PE reduction in crime. This is offset though by the increase in average earnings and therefore return to crime. This latter effect is absent from the analytical example.

[55] These are the same we used in the comparative steady state analysis in Section 5.1.

in the size of the average inter-vivos transfer and a increase in the share of individuals born with zero wealth. In the latter cases the increases in welfare are more than 0.5 percentage points larger than in the benchmark calibration, while the victimization rates drop by more. As expected, the subsidy is more effective in a context in which people are more constrained by initial resources.

Borrowing limit. We reparameterize the benchmark economy to allow for the borrowing limit to be much looser (by a factor of 5) than the one used in our main parameterization. This corresponds to a borrowing limit which is more than 1.5 times the yearly average labor earnings in the economy. As one would expect, the drop in the victimization rate and the welfare gains are smaller, reflecting the reduced selection on wealth, but still sizeable.[56]

Correlation of productive ability and crime fixed effects. Changing $\rho_{\theta\chi}$, the correlation between ability and crime fixed effect, by plus or minus 50% (relative to 0.15 in the main calibration) does not imply substantial changes in crime responses and welfare. This finding suggests that gains associated to the subsidy policy do not hinge on small differences in the composition of the pool of criminals.

Direct cost of high school and college. We also experiment with changing the direct cost of schooling. We, in turn, reduce and increase college tuitions by 50%. This does not change the results in any noticeable way. We also experiment with doubling the HS cost and setting them to zero, again with no noticeable change in results.

7 Conclusions

We develop and calibrate a structural, life-cycle model with heterogeneous agents and optimal education, crime and saving decisions. We use the model to study property crime and compare two alternative sets of policies: subsidies for high school completion and increases in prison sentences. We find that, given the same target in crime reduction, a subsidy to high school completion has large efficiency and welfare gains which are absent in the case of increases in prison sentences.

[56] To understand why such a dramatic improvement in access to credit does not have more dramatic effects, one needs to keep in mind that the subjective discount factor has been recalibrated to keep the wealth-income ratio at its targeted value of 2.7 per cent.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2012): "Equilibrium Effects of Education Policies," Mimeo, University College London.

Attanasio, O. (1999): "Consumption," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1B. North Holland, Amsterdam.

Barillas, F., and J. Fernandez-Villaverde (2007): "A Generalization of the Endogenous Grid Method," *Journal of Economic Dynamics and Control*, 31, 2698–2712.

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

Broidy, L. M., D. S. Nagin, R. E. Tremblay, J. Bates, B. Brame, K. A. Dodge, D. Fergusson, J. L. Horwood, R. Loeber, R. Laird, D. R. Lynam, T. E. Moffitt, G. S. Pettit, and F. Vitaro (2003): "Developmental Trajectories of Childhood Disruptive Behaviors and Adolescent Delinquency: A Six-Site, Cross-National Study," *Developmental Psychology*, 39, 222–245.

Card, D., and T. Lemieux (2001): "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, 116(2), 705–746.

Carroll, C. D. (2006): "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems," *Economics Letters*, 91, 312–320.

Clausen, A., and C. Strub (2012): "Envelope Theorems for Non-Smooth and Non-Concave Optimization," Mimeo, University of Pennsylvania.

Cooley, T. F. (1995): *Frontiers of Business Cycle Research*. Princeton, N. J.: Princeton University Press.

Cozzi, M. (2006): "Hard Drug Addiction, Drug Violations and Property Crime in the US," Mimeo, Queen's University.

Cunha, F., J. Heckman, and S. Navarro (2004): "Counterfactual Analysis of Inequality and Social Mobility," Mimeo, University of Chicago.

De Nardi, M. (2004): "Wealth Inequality, Intergenerational Links and Estate Taxation," *res*, 71, 743–768.

DEARDEN, L., C. EMMERSON, AND C. MEGHIR (2009): "Conditional Cash Transfers and School Dropout Rates," *Journal of Human Resources*, 44, 827–857.

DOMEIJ, D., AND J. HEATHCOTE (2003): "On The Distributional Effects Of Reducing Capital Taxes," Mimeo.

DONOHUE, J., AND P. SIEGELMAN (2004): "Allocating Resources among Prisons and Social Programs in the Battle against Crime," *Journal of Human Resources*, 39, 958–979.

FELLA, G. (2011): "A Generalized Endogenous Grid Method for Non-Concave Problems," Working Paper 677, Queen Mary University of London.

FREEMAN, R. (1999): "The Economics of Crime," in *Handbook of Labor Economics*, ed. by D. Card, and O. Ashenfelter, vol. 3C, chap. 52. Elsevier Science Publishers.

GOLDIN, C., AND L. F. KATZ (2010): *The Race between Education and Technology.* Harvard University Press.

GOULD, E., B. WEINBERG, AND D. MUSTARD (2002): "Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997," *Review of Economics and Statistics*, 84(1), 45–61.

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

HAHN, A., T. LEAVITT, AND P. AARON (1994): *Evaluation of the Quantum Opportunities Program: Did the Program Work?* Brandeis University, Heller Graduate School, Waltham, MA.

HEATHCOTE, J., K. STORESLETTEN, AND G. VIOLANTE (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *jpe*, 118, 681–722.

HECKMAN, J., L. LOCHNER, AND C. TABER (1998): ""Explaining Rising Wage Inequality: Explanations With A Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents," *Review of Economic Dynamics*, 1, 1–58.

HECKMAN, J. J., L. J. LOCHNER, AND P. E. TODD (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of* *the Economics of Education*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 7, pp. 307–458. Elsevier.

Imrohoroglu, A., A. Merlo, and P. Rupert (2000): "On the Political Economy of Welfare Redistribution and Crime," *International Economic Review*, 41, 1–25.

——— (2004): "What Accounts for the Decline in Crime," *International Economic Review*, 45, 707–729.

Kane, T. J. (2003): "A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going," NBER Working Papers 9703, National Bureau of Economic Research.

Kaplan, G., and G. L. Violante (2010): "How Much Consumption Insurance beyond Self-Insurance?," *American Economic Journal: Macroeconomics*, 2, 53–87.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107(1), 35–78.

Lee, D., and K. I. Wolpin (2004): "Intersectoral Labor Mobility and the Growth of the Service Sector," PIER working paper 04-36.

Levitt, S. (1996): "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation," *Quarterly Journal of Economics*, 111, 319–352.

——— (1997): "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 87, 270–290.

Lewis, G. L. (1989): "Trends in Student Aid: 1963-64 to 1988-89," *Research in Higher Education*, 30, 547–561.

Lochner, L. (2004): "Education, work and crime: a human capital approach," *International Economic Review*, 45(3), 811–843.

Lochner, L. J., and E. Moretti (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94(1).

Machin, S., and C. Meghir (2004): "Crime and Economic Incentives," *Journal of Human Resources*, 39, 958–979.

Merlo, A., and K. I. Wolpin (2009): "The Transition from School to Jail: Youth Crime and High School Completion Among Black Males," PIER Working Paper No. 09-002.

Nagin, D. S., D. P. Farrington, and T. E. Moffitt (1995): "Life-Course Trajectories of Different Types of Offenders," *Criminology*, 33, 111–139.

Nagin, D. S., and K. C. Land (1993): "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model," *Criminology*, 31, 327–362.

Nagin, D. S., and R. Paternoster (1991): "On the Relationship of Past to Future Participation in Delinquency," *Criminology*, 29, 163–189.

NCHS (1997): *US Decennial Life Tables for 1989-1991*. Hyattsville:MD.

Nelsen, R. B. (2006): *An Introduction to Copula*. Springer.

Pastore, A. L., and K. Maguire (eds.) (various years): *Sourcebook of criminal justice statistics*. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC.

Raphael, S., and J. Ludwig (2003): "Prison Sentence Enhancements: The Case Project Exile," in *Evaluating Gun Policy: Effects on Crime and Violence*, ed. by J. Ludwing, and P. Cook, chap. Chapter 7. Brookings Institutions, Washington, DC.

Taggart, R. (1995): *Quantum Opportunities Program*. Opportunities Industrialization Centers of America, Philadelphia.

Wolff, E. N. (2000): "Recent trends in wealth ownership, 1983-1998," working paper.

A Appendix

A.1 Analytical example: a simple two period model

Let $v^e(\theta) = \log(w^e\theta) - \log\bar{c}$ denote the opportunity cost of crime for a worker of type (e,θ). Equation (17) can be rewritten as

$$U_2^e(\theta) = \log\bar{c} + \max\{v^e(\theta), u + (1-\pi)v^e(\theta)\}.$$

and the reservation rule (18) as $u_r^e(\theta) = \pi v^e(\theta)$. Integrating over $[-a, a]$ using the reservation rule, yields

$$\begin{aligned}
\mathbb{E}U_2^e(\theta) =& \log\bar{c} + \frac{1}{2a}\left(\int_{-a}^{\pi v^e(\theta)} v^e(\theta)du + \int_{\pi v^e(\theta)}^{a} [(1-\pi)\, v^e(\theta) + u]du\right) \\
=& \log\bar{c} + v^e(\theta) - \frac{\pi v^e(\theta)[a - \pi v^e(\theta)]}{2a} + \frac{a^2 - \pi^2 v^e(\theta)}{4a} \\
=& \log\bar{c} + \left(1 - \frac{\pi}{2}\right) v^e(\theta) + \frac{a}{4} + \frac{\pi^2 v^e(\theta)}{4a}.
\end{aligned}$$

Assuming that $\frac{\pi^2}{2} \simeq 0$ one can disregard the last term and, replacing in the expressions for $U_1^e(\theta)$ and setting $U_1^H(\theta) = U_1^L(\theta)$, obtain an expression for the cutoff ability θ^*

$$\log\theta^* = d - sub - \left(1 - \frac{\pi}{2}\right)\log\frac{w^H}{w^L}$$

as a function of the net fixed cost of education $d - sub$ and the skill premium.

Replacing for the skill premium using (20) yields the (general) equilibrium value of θ^* in implicit form

$$\log\theta^* = d - sub - \left(1 - \frac{\pi}{2}\right)\left[\alpha + (1-\varrho)\log\left(\frac{\theta^{*2}}{1-\theta^{*2}}\right)\right].$$

The above expression implies that the general equilibrium response of the equilibrium dropout share θ^* to a high school subsidy is given by

$$\left[\frac{1 + (2-\pi)\,(1-\varrho)}{\theta^*} + \frac{(2-\pi)\,(1-\varrho)\,\theta^*}{1-\theta^{*2}}\right] d\theta^* = -dsub$$

which can be rearranged as

$$\frac{d\theta^*}{dsub} = -\theta^* \frac{1-\theta^{*2}}{1-\theta^{*2} + (2-\pi)(1-\varrho)}, \tag{26}$$

equation (21) in the main text.

To obtain the probability that a worker of type (e, θ) is a criminal note that the reservation rule (18) implies

$$\Pr\{u \geqslant \pi(\log w^e + \log\theta - \log\bar{c})\} = \frac{1}{2a}\left[a - \pi(\log w^e + \log\theta - \log\bar{c})\right].$$

Integrating over ability, yields the measures of criminals within each education group,

$$\begin{aligned}
&\#\text{ of criminals of edu=L:} \quad \textstyle\int_0^{\theta^*} \frac{1}{2a}\left[a - \pi\left(\log w^L + \log\theta - \log\bar{c}\right)\right] d\theta,\\
&\#\text{ of criminals of edu=H:} \quad \textstyle\int_{\theta^*}^{1} \frac{1}{2a}\left[a - \pi\left(\log w^H + \log\theta - \log\bar{c}\right)\right] d\theta.
\end{aligned}$$

Adding the two, one obtains that aggregate measure of criminals

$$C = \frac{1}{2a}\left[a - \pi(\mathbb{E}\log\theta - \log\bar{c}) + \pi\theta^*\log\left(\frac{w^H}{w^L}\right) - \pi\log w^H\right],$$

which is equation (21) in the main text. Finally, replacing for the skill premium using (20) one obtains

$$C = \frac{1}{2a}\left[a - \pi(E\log\theta - \log\bar{c}) + \pi\theta^*\left\{\alpha + (1-\varrho)\log\left(\frac{\theta^{*2}}{1-\theta^{*2}}\right)\right\} - \pi\log w^H\right].$$

One can use the above expression to show that $\frac{\partial C}{\partial\theta^*}$ is the sum of three terms, that is:

$$\frac{\partial C}{\partial\theta^*} = \frac{\pi}{2a}\left\{\left[\alpha + (1-\varrho)\log\left(\frac{\theta^{*2}}{1-\theta^{*2}}\right)\right] + \frac{2(1-\varrho)}{1-\theta^{*2}} - \frac{\partial\log\left(w^H\right)}{\partial\theta^*}\right\}.$$

The first addendum in the brace is the log skill premium and is unambiguously positive as long as the skill premium is positive. The second addendum is also positive. As for the third addendum, it follows from $w^H = (1-s)H_H^{\varrho-1}H^{1-\varrho}$ that it is given by

$$\frac{\partial\log\left(w^H\right)}{\partial\theta^*} = 2(1-\varrho)\theta^* + \frac{\left(\frac{1-\varrho}{\varrho}\right)\left[2s\varrho\theta^{*(2\varrho-1)} + (1-s)\varrho\left(1-\theta^{*2}\right)^{\varrho-1}(-2\theta^*)\right]}{\left[s\frac{\theta^{*2\varrho}}{2} + (1-s)\frac{\left(1-\theta^{*2}\right)^{\varrho}}{2}\right]}$$

or, after rearranging,

$$\frac{\partial\log\left(w^H\right)}{\partial\theta^*} = 2(1-\varrho)\theta^*\left\{1 + \frac{\left[s\theta^{*2(\varrho-1)} - (1-s)\left(1-\theta^{*2}\right)^{\varrho-1}\right]}{\left[s\frac{\theta^{*2\varrho}}{2} + (1-s)\frac{\left(1-\theta^{*2}\right)^{\varrho}}{2}\right]}\right\}.$$

A.2 Stationary Equilibrium

Let $z_j^x \in Z_j^x$ denote the state implicit in the recursive representation of the problem for an individual of age j and type x, where x can take value s (student), n (worker out of jail) and p (worker in jail).

For a given set of government policies $\{pen, G, sub^e, t_l, t_k\}$, school fees $f^e$and apprehension probability π_p, a *stationary recursive equilibrium* is a collection of (i) policy functions for consumption $\{c_j^s(z_j^s), c_j^s(z_j^s)\}$, saving $a_{j+1}^x(z_j^x)$, bequests b, education $\{i_H^s(z_0^s), i_C^s(z_{j_H}^s)\}$ and crime $\{\tau_j^n(z_j^n)\}$; (ii) value functions $\{V_j^x(z_j^x)\}$; decision rules $\{K, H_L, H_H, H_C\}$ for firms; (iv) prices $\left\{r, w^L, w^H, w^C\right\}$; (v) a victimization rate π_v; (vi) an average labor income $\overline{wh}$; (vii) time-invariant measures $\{\mu_j^s, \mu_j^n\}$ $\Gamma(a)$ that satisfy the following conditions.

1. Given prices$\left\{r, w^L, w^H, w^C\right\}$:
   - for $x = s, n$ the decision rules $\{c_j^x(z_j^x), a_{j+1}^x(z_j^x)\}$ and the value functions $V_j^x(z_j^x)$ solve respectively equations (9)-(11) for $x = s$, equation (12) for $x = n$ and $j < j_r, j \neq j_b$, equation (14) for $x = n$ and $j = j_b$, equation (16) for $j \geq j_r$;
   - the decision rule $a_{j+1}^p(z_j^x)$ satisfies equation (7) with $i^p = 1$ and the associate value function $V_j^p(z_j^p)$ solves equation (13) if $j \neq j_b$ and (15) if $j = j_b$.
   - the decision rule b solves equation (14);
   - the education decisions $\{i_L^s(z_0^s), i_H^s(z_{j_H}^s)\}$ solve equations (8) and (11);
   - the crime decision $\tau_j(z_j^n)$ solves equation (12).
2. Given prices$\left\{r, w^L, w^H, w^C\right\}$, input demands $\{K, H_L, H_H, H_C\}$ maximize profits for the representative firm

   $$r = (1 - t_k)(F_K - \delta)$$

   and

   $$w^e = (1 - t_l)F_{H_e}, \text{ for } e \in \{L, H, C\}.$$

3. The asset market clears

   $$K = \sum_{j,x} \int_{Z_j^x} a_{j+1}^x(z_j^x) d\mu_j^x$$

   .

4. The labor markets for each educational level clear[57]

$$H_e = \sum_{j<j_r} \int_{\{z_j^n : e=i\}} h_j(\theta, e)(1 - \pi_p \tau(z_j^n)) d\mu_j^n, \text{ for } i \in \{0,1,2\}.$$

where the supply of labor on the right hand side of the above equation is made up only of individuals out of jail. These are a fraction $(1 - \pi_p \tau(z_j^n))$ of workers in their age group, in stationary equilibrium.

5. The government budget is balanced

$$G + E + PRIS + PENS = \frac{t_k}{1-t_k} rK + \frac{t_l}{1-t_l} \sum_e w^e H_e$$

Total government outlays on the left hand side of the above equation are the sum of exogenous wasteful expenditure G, education subsidies $E = \sum_{j,i} \int_{\{z_j^s : e=i\}} sub^i \, d\mu_j^s$, for $i = \{L, H, C\}$, aggregate prison expenditure[58] $PRIS = \sum_{j<j_r} \int_{Z_j^n} m \, \pi_p \tau(z_j^n)) d\mu_j^n$ and aggregate pension expenditure $PENS = \sum_{j \geq j_r} \int_{Z_j^n} pen \, d\mu_j^n$.

6. The victimization rate coincides with the crime rate

$$\pi_v = \left(\sum_{j<j_r} \int_{Z_j^n} (1 - \pi_p \tau(z_j^n)) d\mu_j^n \right)^{-1} \sum_{j<j_r} \int_{Z_j^n} \tau_j(z_j^n) d\mu_j^n,$$

and equals the total number of crimes divided by the total number of workers out of jail.

7. The average disposable labor income satisfies

$$\overline{wh} = \left(\sum_{j<j_r} \int_{Z_j^n} (1 - \pi_p \tau(z_j^n)) d\mu_j^n \right)^{-1} \sum_e w^e H_e.$$

8. The distribution of wealth at birth $\Gamma(a_0)$ equals the distribution of bequests

$$\Gamma(a_0) = \int_{\{z_{j_b}^n : b(z_{j_b}^n) \leq a_0\}} d\mu_{j_b}^n + \mathbb{I}_{a_0=0} \tau_{j_b}(z_{j_b}^n) \mu_{j_b}^n$$

9. The vector of measures $\mu = \{\mu_0^s, ..., \mu_J^s; \mu_0^n, ..., \mu_J^n\}$ is the fixed point of $\mu(Z) = Q(Z, \mu)$ where Z is the generic subset of the Borel sigma algebra $\mathfrak{B}_Z$ defined over

[57] By Walras law, market clearing on all factor markets ensures that the goods market clears.

[58] In stationary equilibrium, the number of convicted felons in each age group equals a fraction $\pi_p \tau(z_j^n))$ of the corresponding number of workers.

the state space $\mathbb{Z} = \prod_{j,x} Z_j^x$, the Cartesian product of all Z_j^x. The mapping $Q(Z, \mu)$ is the transition function associated with the individual decisions, the law of motion for the shocks $\{\chi, \theta, \upsilon, i^p, \varepsilon_j^e\}$ and the survival probabilities $\{\lambda_j\}$.

A.3 Computation and calibration

Let $Z = \{\xi, \nu_1, \nu_2, \nu_3, \underline{a}, \underline{\chi}, \alpha, \beta, \rho_{\theta\chi}, \kappa, \bar{c}\}$ denote the set of calibrated parameters other than the utility cost of studying parameters $\{\psi^H(\theta), \psi^C(\theta)\}$. Given a guess for Z, $\{\psi^H(\theta), \psi^C(\theta)\}$ and the vector of equilibrium prices $\{r, w^L, w^H, w^C\}$ we calibrate the model in the following way.

1. We solve for the consumer decisions rules and value function and the representative firm factor demand functions.

2. We simulate the model up to the age of the college choice and solve for the values of $\{\psi^H(\theta), \psi^C(\theta)\}$ that match the enrolment rates in the data. Using the new values as our new guess, we simulate again the economy up to the age of college choice and iterate on this procedure until the values of $\{\psi^H(\theta), \psi^C(\theta)\}$ converge.

3. We simulate the model at the remaining ages and compute the aggregate factor supplies. The compare the marginal products of the four factors to our guess for their prices. If the two differ by more than the specified tolerance, we adjust the guess for prices and solve again the problem, starting from point 1. until convergence (market clearing).

4. When factor prices have converged, we evaluate the loss function – the sum of squared deviations of the model from the data calibration moments – at the simulated model moments. We use multi-dimensional Newton-Raphson methods to update the guess on Z and continue to iterate starting from point 1. above until convergence.

Concerning point 1. the decision rules and value functions point are computed using a generalized version of the endogenous grid method developed in Fella (2011). The method extends the original idea of Carroll (2006) to environments with non-convex choice sets.[59]

[59] Barillas and Fernandez-Villaverde (2007) extend the endogenous grid method to perform value function iteration in models with more than one control variable, but with a convex choice set.

While the reader is referred to Fella (2011) for the details, we include here a brief sketch of the algorithm in the context of a simple problem.

Consider an agent with a two-period lifetime who derives intra-period utility $u(c, d)$ from consuming quantity c of a continuous good and quantity $d \in D = \{0, 1\}$ of a discrete good. The utility function satisfies the usual regularity conditions and, for simplicity, the Inada condition $u'(0, \cdot) = +\infty$. The relative price of the two goods is one. The agent has an initial endowment a of the continuous good. Both the (net) rate of return on storage and the agent subjective discount rate equal zero. There is no borrowing.

The agent's problem in recursive form is

$$\begin{aligned} v(a) &= \max_{a' \in [0,a], d \in D} u(a - a' - d, d) + v'(a') \\ v'(a') &= \max_{a'' \in [0,a], d' \in D} u(a' - a'' - d', d') \\ &\quad a \text{ given.} \end{aligned} \tag{27}$$

It follows that $v'(a') = u(a' - \hat{d}'(a'), \hat{d}'(a'))$ with $\hat{d}'(a') = \arg\max_{d' \in D} u(a' - d', d')$. The non-convexity of D, implies that, to the extent that $\hat{d}'(a')$ is not a constant, $v'(a')$ is neither concave nor differentiable and neither is the maximand and on the right hand side of (27).

Yet, Theorem 2 in Clausen and Strub (2012) implies that if, for given a, $(\hat{a}', \hat{d})$ is a maximum for (27) and $\hat{a}'$ is internal then $\hat{a}'$ satisfies the Euler equation

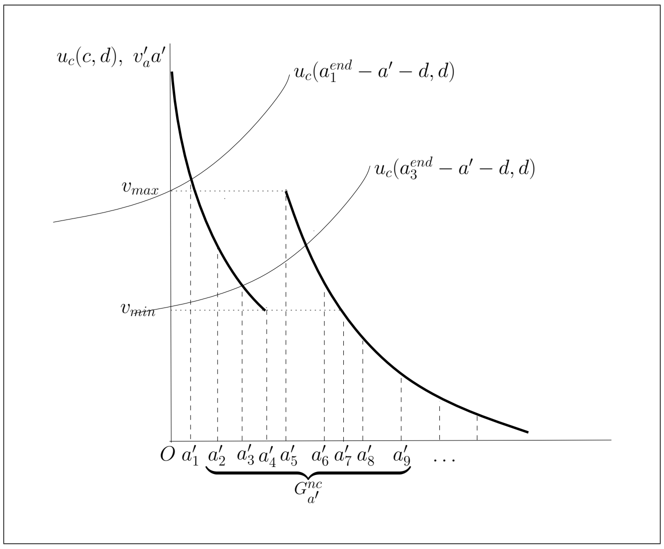
$$u_c(a - \hat{a}' - \hat{d}, \hat{d}) = v'_a(\hat{a}'), \tag{EE}$$

as $v'_a(a')$ can jump up but not down.[60]

Figure A plots the right and left hand sides of equation EE as a function of a' for a given value of d, The left hand side is plotted for two possible values of initial assets a. For given a, the intersection of the two curves is a candidate solution for the saving correspondence $a'(a|d)$ *conditional* on the given value of d. The contribution of Fella (2011) concerns how to solve for this "conditional" saving correspondence $a'(a|d)$.

[60]This implies that the value of the Euler equation jumps up at discontinuities of $v'(a')$. Therefore a maximum cannot be located at a discontinuity.

Figure A: Solving for the conditional policy correspondence



In the standard approach, one fixes values for the endogenous state variable a at the beginning of the period and solves the Euler equation forward for the associated values of end-of-period wealth a'. Carroll's (2006) endogenous grid method (EGM), instead, fixes an ordered grid $G_{a'} = \{a'_1, a'_2, \dots, a'_m\}$ for *end-of-period* assets a' and solvea for the value of *initial wealth* a_i^{end} that satisfies EE for each $a'_i \in G_{a'}$. for each $a'_i \in G_{a'}$.[61] This approach is substantially faster as the Euler equation is often linear in consumption, hence in a, but non-linear (and in our case not even continuous) in a'.

Since, given the non-concavity of the problem, a local maximum is not necessary a global one, the algorithm modifies the standard EGM in the following way. First, it partitions the set of grid points for future assets $G_{a'}$ into a *non-concave region* $G^{nc}_{a'}$ in which the Euler equation is not sufficient for a global maximum for a' and its set complement. In terms of Figure A, given the grid $G_{a'}$ and the derivative of the continuation value $v_a(a')$ it determines the *non-concave region* $G^{nc}_{a'}$ as the set of grid points for which $v'_a(a') \in (v_{min}, v_{max})$.[62] Secondly, for all a'_i in the non-concave region, the algorithm supplements EGM with a global maximization step.

More formally, given $G_{a'}$, $v'_a(a')$ for $a' \in G_{a'}$ and d

[61] In terms of of Figure A, at the grid point a'_1, for example, the EGM solves for the value of initial assets a_1^{end} associated with the unique element of the family of upward sloping curves, indexed by initial wealth a, that intersects $v'_a(a')$ at point a'_1.

[62] In Figure A, $G^{nc}_{a'} = \{a'_2, \dots, a'_6\}$.

1. Determine the non-concave region $G_{a'}^{nc}$. Initialize the counters $i = 1$ and $l = 1$
2. Solve EE for a_i^{end} given a_i' using EGM.
3. If $a_i' \in G_{a'}^{nc}$ then
   - find the maximizer of the discretized maximand for $a = a_i^{end}$; i.e. solve for

$$a_g' = \arg \max_{a' \in G_{a'}^{nc}} u(a_i^{end} - a' - d, d) + v'(a').$$

   - if $a_g' \neq a_i'$, a_i' is not a global maximum. Move to the next grid point – $i = i + 1$ – and go to 2.
4. Store the solution pair (a_i^{end}, a_i') as $(a_{i_l}^{end}, a_{i_l}') = (a_i^{end}, a_i')$. As long as a' is not the last grid point, set $i = i + 1, l = l + 1$ and go to 2.
5. Having solved for the conditional saving correspondence $\{a_{i_l}^{end}, a_{i_l}'\}$ on the endogenous collocation points $\{a_{i_l}^{end}\}$ solve for the conditional value function given d

$$v_{i_l} = u(a_{il}^{end} - a_{il}' - d, d) + v'(a')$$

6. Evaluate interpolating functions through (a_{il}^{end}, a_{il}') and (a_{il}^{end}, v_{il}) at $a \in G_{a'}$ to obtain the conditional policy and value functions $a'(a|d)$ and $v(a|d)$ on the original grid $G_{a'}$.
7. Maximize $v(a|d)$ over d to obtain $d(a)$ and $v(a)$.

In a longer (possibly infinite) horizon case, having obtained $v(a)$ one would compute its partial derivative $v_a(a)$ and would work backwards.

Fella (2011) compares the accuracy and speed of the method to that of discretized value function iteration (VFI) – the most commonly chosen algorithm for non-concave, non-differentiable problems – using a saving problem with a discrete durable and a continuous non-durable choice. The discrete non-durable choice can take seven values, which implies a number of potential discontinuities larger than in the current model. He finds that the modified EGM algorithm has an accuracy, measured by the average Euler error (in base 10 log points) over a simulated history, in excess of -5 already with only 200 grid points for the continuous wealth variable. This is more than twice the accuracy of VFI[63]

[63]The average Euler error, rather than the supremum of the Euler errors, is the sensible accuracy measure in a model with discontinuities in the policy function, since, no matter how large the number of

for the same number of grid points.[64]

grid points, the probability of interpolating across a discontinuity goes to one as the length of a history increases. The Euler error when interpolating across the discontinuity is determined by the size of the jump in the function.

[64]In fact, the modified EGM with 200 grid points is still two orders of magnitudes more accurate, and 70 times faster, than VFI with 1000 grid points.

Table 1: Value of assigned parameters in the benchmark

Parameter	Value	Moment to Match
$\{\lambda_j\}$	See text	Survival rates (US Life Tables)
$\bar{j}$	79	Real-world age of 95
j_b	45	Bequest age
j_r	50	Maximum years of working life
ϕ	0.35	Capital share of output (NIPA)
s_e	See tab. 3	Human capital shares
ϱ	0.677	Elasticity of substitution
δ	6.5%	Depreciation rate (NIPA)
t_l	27%	Labor income tax rate
t_K	40%	Capital income tax rate
pen	1,980$	16% pension replacement rate
m	8,300$	Avg. cost per per prisoner
j_H	2	Post-compulsory high school duration
j_C	4	4-year college duration
f^H	124$	Direct cost of high school
f^C	1,200$	Direct cost of college
γ^e	See tab. 4	Earnings ability gradient
$\zeta^e(j)$	See tab. 5	Earnings life cycle profile
ρ^e, σ^e	See tab. 6	Earnings residual dynamics
RRA coefficient	1.5	Micro estimates
π_p	.057	Probability of conviction
Sentence length	19 months	Average completed sentence

Table 2: Value of calibrated parameters and targeted moments.

Parameter	Value	Moment to Match	Data	Model
ξ	0.967	Wealth-income ratio excluding top 5%	2.7	2.74
ν_1	1.36	Average inter-vivos transfer	15,200$	14,900$
ν_2	3.38	% of young receiving no inter-vivos transfers	9	9
ν_3	0.91	Coefficient of variation of inter-vivos transfers	0.75	0.75
$\underline{a}$	-3,960$	% of households with net worth ≤ 0	15	15
$\psi^H(\theta), \psi^C(\theta)^a$		Enrolment rates by ability bin (see Table 7)		
κ	0.059	Average loss from property crime	728$	728$
$\overline{c}$	3,900$	Victimization rate (%)	5.6	5.6
$\rho_{\theta\chi}$	0.12	% of HS dropouts among prisoners	53	48
Regression:[b]				
–		Regression constant	25.2	22.3
		s.e.	(11.7)	
$\underline{\chi}$	-.84	Education coefficient	-6.3	-4.5
		s.e.	(2.3)	
α	0.72	AFQT89 pct. (1-99) coefficient	0.04	0.04
		s.e.	(0.03)	
β	1.15	Age coefficient	-0.71	-0.71
		s.e.	(0.59)	

Notes:

[a]Values for $\psi^H(\theta), \psi^C(\theta)$ available upon request.

[b]Unit of observation is individuals aged 18-23 in the model data and males aged 18-23 in 1980 in the NLSY79. The dependent variable is a dummy equal to one if the individual participated in property crime. All coefficient estimates are multiplied by 100.

Table 3: Production shares of different types of human capital in the years 1980 and 2000. Source: Gallipoli, Meghir and Violante (2012).

Year	High School Dropouts	High School	College
1980	0.22	0.41	0.37
2000	0.16	0.39	0.45

Table 4: Estimated ability gradient. Source: Gallipoli, Meghir and Violante (2012).

Education group	Gradient (Std. err.)
High School Dropouts	.36
	(.06)
High School	.54
	(.03)
College	.89
	(.09)

Table 5: Age polynomials' coefficients. Dependent variable: real log hourly earnings ($1992). Source: Gallipoli, Meghir and Violante (2012).

	Less Than HS Coefficient (Std. err.)	High School Coefficient (Std. err.)	College Coefficient (Std. err.)
age	0.26	0.41	0.67
	(0.133)	(0.058)	(0.101)
age^2	-0.01	-.013	-.021
	(0.005)	(0.002)	(0.004)
age^3	1.5e-4	1.e-4	3.e-4
	(8.3e-5)	(4.e-5)	(6.e-5)
age^4	-9.7e-7	-1.e-6	-1.6e-6
	(4.9e-7)	(2.e-7)	(3.7e-7)
Intercept	-0.507	-2.23	5.12
	(1.232)	(0.533)	(0.967)

Table 6: Estimated autoregressive coefficient and variance for the persistent shocks to wages, by education group. Source: Gallipoli, Meghir and Violante (2012).

	H.S. dropouts	H.S. graduates	Coll. graduates
ρ^e	0.936	0.950	0.945
σ^e	0.016	0.010	0.014

Table 7: Shares of workers in different education groups by ability (AFQT89). Values are grossed-up to replicate aggregate education shares observed for workers between 1977 and 1983 in CPS March supplement.

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Aggregate
HSD	86	25	11	3	1	25
HSG	13	70	78	73	54	57
CG	1	5	11	24	45	17

Table 8: Model implications

Statistic	Data	Model
Effect of education on incarceration[a]		
HS school graduation dummy	-0.49	-0.37
S.E.	(0.39)	
Effect of education on crime participation[b]		
HS school graduation dummy	-5.63	-4.07
S.E.	(2.25)	
Enrollment response		
High school enrollment (pct. points)[c]	6.7	6.6
S.E.	(1.7)	
College enrollment (pct. points)[d]	3 to 9	3.6/4.9
Property crime and education shares in 2000		
Overall effect: all changes together		
Crime rate (%)	3.6	3.6
Share of high school graduates (%)	60	59
Share of college graduates (%)	26	27.5
Changing: Human capital shares (skill bias):		
Crime rate (%)	-	8.0
Share of high school graduates (%)	-	65
Share of college graduates (%)	-	32
Changing: Expected punishment:		
Crime rate (%)	-	2.1
Changing: Income variance:		
Crime rate (%)	-	5.8
Changing: College tuitions:		
Crime rate (%)	-	5.7
Share of high school graduates (%)	-	59
Share of college graduates (%)	-	15.5
Changing: Demographics:		
Crime rate (%)	-	5.3

Notes:

[a]The data numbers are obtained by estimating the equation in Table 9, column 4 in Lochner and Moretti (2004) but for year 1980 alone. Coefficient multiplied by 100.

[b]The data numbers are obtained by estimating the equation in Table 12, column 4 in Lochner and Moretti (2004) for the pooled sample. Coefficient multiplied by 100.

[c]The data numbers are taken from Dearden, Emmerson, and Meghir (2009). Estimated response of high school enrollment to a high school subsidy equal to 20 per cent of the average post-tax earnings of dropouts aged 16.

[d]The data numbers are taken from Kane (2003). Estimated response of college enrollment to a change in college tuition fees equal to $1000 in 2001. The model numbers are respectively for a decrease and an increase in the fees.

Table 9: Differences with respect to benchmark (absolute changes) in shares of workers in different education groups by ability (IQ test) bin, given a non-means tested high school subsidy (partial equilibrium)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Average
HSD	-17	-14	-11	-3	-1	-9
HSG	17	14	9	-2	-7	6
CG	0	0	2	5	8	3

Table 10: Differences with respect to benchmark (absolute changes) in shares of workers in different education groups by ability (AFQT89) bin, given a non means-tested high school subsidy (general equilibrium).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Average
HSD	4	1	-8	-3	-1	-1
HSG	-4	0	11	1	-5	1
CG	0	-1	-3	1	6	0

Table 11: Subsidy and prison experiments. Subsidy as % of average labour income.

	Benchmark	HS Subsidy PE	HS Subsidy GE		Prison	
	(1)	(2)	(3)	(4)	(5)	(6)
HS subsidy	-	8.8	8.8	17.6	-	-
Prison sentence (months)	19	19	19	19	20.0	20.5
Crime victimization (%)	5.62	5.44	5.19	4.96	5.19	4.96
Arrest rate HSD (‱)	5.93	6.63	5.31	4.87	5.55	5.31
Arrest rate HSG (‱)	2.73	2.96	2.65	2.58	2.50	2.41
HSD share of criminals (%)	48	36	45	42	49	49
Output	100.0	-	101.1	101.8	99.9	99.9
Agg. Consumption	100.0	-	101.3	102.2	99.9	100.0
Welfare	100.0	103.4	103.5	106.4	99.9	99.9
Prison expenditure†	0.27	0.26	0.25	0.25	0.26	0.26
Subsidy + prison exp.†	0.27	0.51	0.48	0.71	0.26	0.26
Price HSD	100.0	-	102.8	104.8	100.0	100.0
Price HSG	100.0	-	100.4	100.8	100.0	100.0
Price CG	100.0	-	99.7	99.4	100.0	100.0

† As a share of aggregate consumption in the benchmark.

Table 12: Results of sensitivity analysis. Each line corresponds to an alternative parametrization and reports the parameter being changed, the re-calibrated benchmark outcomes and the results of the 8.8% High School subsidy experiment in G.E. (for technology parameters also P.E. results are reported). Two outcomes are reported: (i) victimization rate; (ii) change in ex-ante welfare (consumption equivalents) expressed as a share of its benchmark value.

	Crime victimization			Welfare	
	Benchmark	PE	GE	PE	GE
High elasticity (5)	5.61	5.26	5.17	103.3	103.1
Low elasticity (2)	5.60	5.67	5.23	103.0	103.3
Labor shares - year 2000	5.60	5.17	4.95	105.7	103.0
Higher average intervivos	5.64		5.26		103.1
Lower average intervivos	5.62		5.09		104.3
Higher % zero initial wlth	5.63		5.09		104.2
Lower % zero initial wlth	5.60		5.23		102.7
Looser borrowing limit	5.56		5.24		102.1
Higher cost of High School	5.60		5.24		103.4
Lower cost of High School	5.60		5.18		103.2
Higher cost of College	5.61		5.16		103.4
Lower cost of College	5.61		5.21		103.3
Higher $\rho_{\theta\chi}$	5.59		5.17		103.5
Lower $\rho_{\theta\chi}$	5.62		5.20		103.5

Figure 1: Yearly arrest rates for ages 18-25 in data and model.

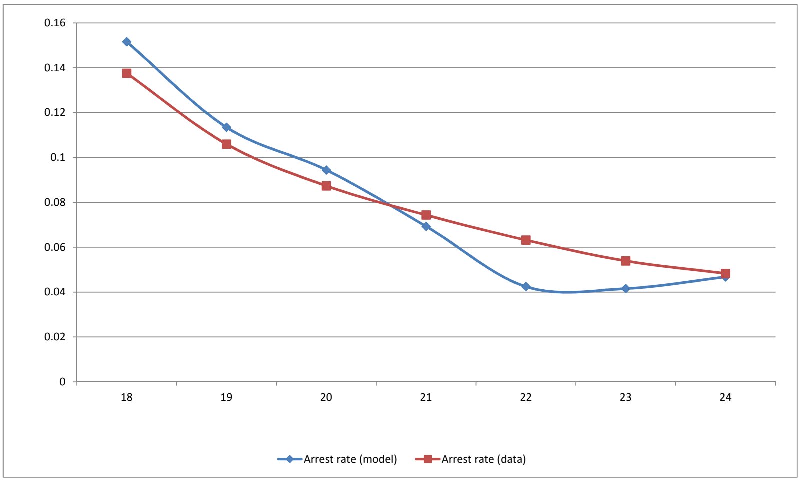


Figure 2: Five-year arrest rates for ages 20-55 in data and model.