Education and Crime over the Life Cycle

Giulio Fella*

Giovanni Gallipoli†

Abstract

We compare two large-scale policy interventions aimed at reducing crime: subsidizing high school completion and increasing the length of prison sentences. To this purpose we use a life-cycle model with endogenous education and crime choices. We apply the model to property crime and calibrate it to U.S. data. 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 also find that general equilibrium effects explain roughly one half of the reduction in crime from subsidizing high school.

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@mail.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 since the early 1980s and currently stands at over 2.2 million inmates.[1] This increase is not the result of an increase in violent crimes, but is instead due to mandatory and longer prison sentencing for non-violent offenders.[2] The average annual cost per prison inmate was 28,900 U.S. dollars in 2008.[3] These figures have prompted a shift of interest, among both academics and policymakers, from tougher sentencing to other forms of intervention.[4]

Educational attainment among the prison population is extremely low. Over seventy percent of U.S. inmates in 1997 did not have a high school degree.[5] In an influential paper Lochner and Moretti (2004) have established a sizable causal effect of education, in particular high school graduation, on crime.[6] On the basis of their estimates Lochner and Moretti calculate that the positive externality in crime reduction generated by an extra high school graduate yields a partial equilibrium net social benefit of about 2,000 U.S. dollars per year, which is between 14% and 26% of the private return to high school graduation.

Given the large monetary and human costs of crime it is important to quantify the relative benefits of policies promoting incarceration vis-a-vis alternative policies such as those aimed at boosting educational attainment. Within the latter category, policies that encourage high school completion have garnered most attention and seem to hold much promise in terms of their impact on crime.[7]

We address this policy question by providing a quantitative analysis of the effect of two

[1]In fact, the U.S. has nearly a quarter of the world prison population against only 5% of the overall population.

[2]See, e.g., Blumstein and Beck (1999) and Mauer (1999).

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

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

[5]See Heckman and LaFontaine (2010). In Italy, more than 75% of inmates in 2001 had not completed high school (Buonanno and Leonida 2006), while incarceration rates among men aged 21-25 in the United Kingdom were more than 8 times higher for those without a secondary education qualification (i.e. dropouts) relative to those with a qualification (Machin, Marie, and Vujić 2011).

[6]Machin, Marie, and Vujić (2011) document a robust, causal impact of education on property crime using UK data.

[7]There is evidence that the most sizable reductions in crime appear to result from completing high school. See Lochner (2011) and references therein.

alternative—economy wide—policies; namely, we compare increasing prison sentences to a high school graduation subsidy. The focus of our analysis is on the long-term effects of policy interventions and, to this purpose, we develop a general equilibrium, life-cycle model with heterogeneous agents, incomplete markets and endogenous education and crime choices. The economy includes a production sector with an aggregate production function that allows for imperfect substitutability among workers with different education levels.

One advantage of a structural model is that it can be used to calculate efficiency *and* welfare gains, or losses, associated with a given policy intervention. By contrast, a reduced-form approach can at best estimate the approximate efficiency gains of a given intervention. If agents are risk averse and incomplete markets cannot insure all realized heterogeneity, then ex ante welfare gains of policies that relax the link between initial conditions and life-cycle outcomes may be substantially larger than the efficiency gains. A robust finding of our analysis is that the former are more than twice as large as the latter.

In addition, the evaluation of the long-run effects of large-scale policy interventions calls for a general equilibrium analysis. Heckman, Lochner, and Taber (1998) have led the way in advocating the use of models that take into account the response of relative prices, and allow agents to anticipate and respond to such adjustment, for policy evaluation.[8] Experimental and quasi-experimental reduced-form techniques have the advantage of not imposing a specific structure at the estimation stage, but they often rely on short-term comparisons and therefore do not identify longer horizon effects. In reality it may take time for the effects of a policy to fully emerge. This is true when interventions alter life cycle choices, even if one abstracts from price feedback.

The model features several sources of heterogeneity. Agents differ ex ante in: 1) innate, observed ability; 2) innate taste for crime;[9] 3) initial wealth. The latter is determined by

[8]Some applications of this approach to policy evaluation are Attanasio, Meghir, and Santiago (2012), Blundell, Dias, and Meghir (2003), Lee (2005), and Lise, Seitz, and Smith (2004). Blundell, Dias, and Meghir (2003) find that the general equilibrium effects of the policy they study are so strong as to reverse the sign of the partial equilibrium effects.

[9]Ability is a set of characteristics that directly affects earnings, whereas the taste for crime 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). Our modeling approach is consistent with a vast literature, especially in criminology, which

inter-generational inter-vivos transfers. On the basis of these differences, individuals in the early stage of their lifetime make education and consumption choices. During their working lifetime, individuals are subject to idiosyncratic and persistent shocks to their labor productivity and choose consumption/saving and 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 U.S. economy in 1980.

Before conducting policy evaluation within the model, we verify its performance along a number of dimensions using data restrictions that are not targeted in the parameterization. We start by assessing the model's ability to generate realistic enrollment responses to high school subsidies of both relatively small and very large magnitudes. To this purpose, we first simulate the effect of a randomized program—evaluated in Dearden, Emmerson, Frayne, and Meghir (2009)—that provided moderate monetary incentives to attend the last two years of high school. Secondly, we simulate the effect of a very large high school graduation subsidy as the one considered in Keane and Wolpin (2000). In both cases the simulated enrolment response is quantitatively comparable to the estimates reported in those papers.

Our second set of experiments aims to assess the ability of the model to generate a realistic response of crime to a change in the share of high school graduates. To this purpose we estimate the effect of high school graduation on the probability of being incarcerated for individuals between 20 and 60 years of age and on crime participation for individuals aged 18-23 using simulated data. These are the counterparts of two of the regressions in Lochner and Moretti (2004). The estimates we obtain for the model data are comparable to the corresponding estimates in Lochner and Moretti (2004). We also verify that the model generates an age profile of arrest rates which is consistent with that observed in the data.

Finally we verify that the model generates a response of college enrollment to college

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.

subsidies in line with empirical estimates and that it is able to account for changes in enrollment and crime rates between 1980 and 2000.

We evaluate two alternative policies: a subsidy towards high school completion and an increase in the prison sentence that generates the same change in the victimization rate as the subsidy policy. The size of the subsidy is 8.8 per cent of average labor earnings for each of the last two year of schooling and is chosen to coincide with the average value of the monetary component in a well-known, small scale program.[10]

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.[11] 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) uncertainty 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,

[10]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.

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

plays an important role in terms of crime reduction. The crime fall associated with the subsidy is only half as large as in general equilibrium.

Our approach is in the tradition of economic models of crime which goes back to Becker's (1968) work. It builds upon the original contribution of Imrohoroglu, Merlo, and Rupert (2004) who are the first to construct a calibrated, structural, general equilibrium model of rational crime choice.[12] 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 also 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.[13] 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 proposes a simplified analytic example to highlight the main mechanisms in the model. Section 3 introduces the general model. Section 4 discusses the numerical parameterization. Section 5 simulates the model, assesses its performance and studies the effect of alternative policies. Section 6 presents additional sensitivity analysis. Section 7 concludes. The Appendix reports more technical details.

[12]Two other dynamic equilibrium models of crime are Imrohoroglu, Merlo, and Rupert (2000) and Cozzi (2006).

[13]In addition to the recent papers by Lochner and Moretti (2004) and Machin, Marie, and Vujić (2011) discussed above, 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 An analytic framework

Before presenting the full life-cycle problem and numerical analysis, we use a two-period model, which can be solved analytically, to overview the main economic forces at work.[14]

The population has measure one and agents differ in their ability θ, distributed uniformly on $[\underline{\theta}, 1]$.[15] An agent's education level e can take two values $\{L, H\}$, 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 H. The utility of leisure is normalized to zero, while the disutility of studying is decreasing in ability and given by $d - \log\theta$.[16]

In period two agents inelastically supply one unit of labor for wage $w^e\theta$. They also draw a random return from criminal activity, yielding additively separable utility ς uniformly distributed on $[-\bar{\varsigma}, \bar{\varsigma}]$.[17] After observing the shock realization, agents decide whether to engage in crime. With probability π, criminals are apprehended, lose their income (from labor and crime) 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 production technology is $Y = [sH_L^{\varrho} + (1-s)H_H^{\varrho}]^{\frac{1}{\varrho}}$, with H_L, H_H the stocks of the two worker types and $\varrho \leq 1$.

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

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

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 expected utility from either enjoying the return from crime and consuming her labor income (if she is not apprehended), or consuming the exogenous amount $\bar{c}$ (if she is apprehended). It follows from equation (1)

[14] See Appendix A.1 for details and derivations.

[15] The lower bound $\underline{\theta}$ is an arbitrarily small positive number. The requirement that $\underline{\theta} > 0$ guarantees that utility is bounded below. In what follows we consider the limit case in which $\underline{\theta}$ tends to zero.

[16] The term d represents the ability-independent component of the cost of studying, including tuitions.

[17] The shock ς may be interpreted as a perturbation to the relative return between legal and non-legal activities. Modelling it as an additive shock keeps the analytic model simple.

that an agent engages in crime if her draw of ς is above the reservation value

$$\varsigma_r^e(\theta) = \frac{\pi}{1-\pi}\left[\log\left(w^e\theta\right) - \log\left(\bar{c}\right)\right], \tag{2}$$

which implies that the probability of engaging in crime is decreasing in the probability of apprehension and, through the opportunity cost in the square bracket, 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 $U_1^L(\theta) = \mathbb{E}U_2^L(\theta)$. For $\pi^2 \simeq 0$ (see Appendix), the marginal ability θ^*, for which an agent is indifferent between studying or not, satisfies

$$\log\theta^* \simeq d - sub - \left(1 - \frac{\pi}{2}\right)\log\frac{w^H}{w^L}. \tag{3}$$

In the limit, as $\underline{\theta}$ goes to zero, θ^* is also the proportion of unskilled workers. Equation (3) can be differentiated to obtain the general equilibrium response of the share θ^* to a high school subsidy

$$\frac{d\theta^*}{dsub} = -\left[\frac{1}{\theta^*} + \left(1 - \frac{\pi}{2}\right)\frac{d\log(w^H/w^L)}{d\theta^*}\right]^{-1}. \tag{4}$$

The equilibrium change in the share of unskilled workers can be decomposed into two effects. The first term in the square bracket is the partial equilibrium response. At constant wages, a reduction in the cost of education reduces the share of unskilled workers θ^*. The second term captures the dampening general equilibrium effect: a fall in θ^* reduces the skill premium and therefore the incentive to acquire education. This second effect is weaker when ϱ increases, and it is absent if $\varrho = 1$. It follows that the equilibrium response of the share of dropouts θ^* to the high school subsidy is negative and its absolute value increases as ϱ approaches 1.[18]

Remark 1 *The equilibrium response $d\theta^*/dsub$ is negative and its absolute value is increasing in ϱ.*

Turning to the crime rate, the measure of workers of type (e, θ) who engage in crime in period 2 coincides with the probability that the realization of the return from criminal activity exceeds the reservation value in equation (2). After integrating over ability and the two education levels, and letting $\underline{\theta}$ approach zero, one obtains the following expression for the total measure of criminals

$$C = \frac{1}{2} + \frac{\pi}{2\bar{\varsigma}(1-\pi)}\left[\theta^* \log\left(\frac{w^H}{w^L}\right) - \log w^H - \left(\int_0^1 \log\theta d\theta - \log\bar{c}\right)\right]. \tag{5}$$

A larger share of unskilled workers affects the crime level through three different channels. The first is a (partial equilibrium) composition effect and is positive. A higher θ^* increases the crime level because unskilled workers have a lower opportunity cost of crime. The second (general equilibrium) effect reinforces the first one. By reducing w^L, an increase in θ^* reduces the opportunity cost of crime for unskilled workers. Finally, to the extent that a higher θ^* increases w^H, it also decreases the crime rate among skilled workers, partly counteracting the other two effects. In the numerical analysis we consistently find that the change in w^H is small relative to the change in w^L. This implies that when $\varrho < 1$ the general equilibrium effect reinforces the partial equilibrium response so that the overall response is positive.

While the sign of $\partial C/\partial\theta^*$ is generally positive, the change in its magnitude as ϱ increases is a priori ambiguous. The partial equilibrium response increases with ϱ, as a larger elasticity implies a larger skill premium; viceversa, the general equilibrium response moves in an opposite direction (see Appendix). We summarize these findings in the following remark.

Remark 2 *The equilibrium response $\partial C/\partial\theta^*$ is generally positive. The effect of a change in ϱ on $\partial C/\partial\theta^*$ is ambiguous.*

[18]One can show that, if the coefficient of relative risk aversion is strictly larger than one, also the partial equilibrium response of enrollment to the subsidy is increasing in ϱ, as a higher ϱ is associated with a larger skill premium. A proof is available upon request.

To sum up, the general equilibrium response of the crime level to an education subsidy is the product of $\partial C/\partial\theta^*$ and $d\theta^*/dsub$. From the two remarks above it follows that this product is negative and a subsidy reduces the equilibrium crime level. Moreover, the effect of ϱ on the magnitude of this product is ambiguous because, as ϱ changes, the respective sizes of $\partial C/\partial\theta^*$ and $d\theta^*/dsub$ may move in opposite directions. In other words, the equilibrium effect of a subsidy on crime may not change much under alternative values of the elasticity of substitution in skill types. Nonetheless this model highlights how the effects of a subsidy policy on crime could in principle change with ϱ, a possibility which we explore in the sensitivity analysis.

3 The model

3.1 Demographics and the life cycle

Demographics: The model has an overlapping generation structure. Let $j \in J = \{0, 1, ..., \bar{j}\}$ denote the age of an individual. In each period, each individual of age $j_b < \bar{j}$ begets an offspring. The conditional probability of surviving from age j to $j+1$ is λ_j with $\lambda_j = 1$ for all $j \leq j_b$ and $\lambda_{\bar{j}} = 0$.[19] 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 $i^s_e \in \{0, 1\}$ 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.[20]

[19] Assuming that $\lambda_j = 1$ for $j \leq j_b$ ensures that each agent has exactly one child.

[20] 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.

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; namely, 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 $\tau_j \in \{0, 1, \dots, \bar{\tau}\}$ to commit[21] and how much to consume/save. If 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. 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.

3.2 Preferences

Individuals have time-separable preferences and discount the future at rate $\xi_j = \hat{\xi}\lambda_j$, where $\hat{\xi}$ is the pure rate of time discount unadjusted for the survival probability. 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

$$U^s(c_j) = u(c_j) + \mathbb{I}_{j<j_H}\psi^H(\theta) + (1 - \mathbb{I}_{j<j_H})\psi^C(\theta). \tag{6}$$

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

[21] 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).

degree $e = H, C$. The disutility of education $\psi^e(\theta)$ changes with the 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.[22] The indicator function $\mathbb{I}_{j<j_H}$ takes value one at age $j < j_H$—when the agent is attending high school—and zero otherwise.

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

$$U^w(c_j, \tau_j, b) = u(c_j) + \mathbb{I}_{\tau_j}\chi + \mathbb{I}_{j=j_b}v(b). \tag{7}$$

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. 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.

The time-invariant direct utility from crime χ plays an important role in the model and should be interpreted as a reduced-form way of capturing the effect of heterogeneity—other than heterogeneity in labor market returns—in the relative utility to crime. The shock 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.[23]

It will become clear that the model has a number of channels, all related to market returns, that generate endogenously the observed heterogeneity in crime behaviour both cross-sectionally and over the lifetime. These mechanisms are: (i) permanent differences

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

[23] In modeling individual heterogeneity above and beyond labor market shocks we resort to a dichotomy between ability θ, which largely overlaps with cognitive skills, and crime preference shocks, which have a direct effect on crime. One might be tempted to relate crime taste shocks to a more general notion of non-cognitive skills (as discussed, for example, by Heckman, Stixrud, and Urzua 2006, and Almlund, Duckworth, Heckman, and Kautz 2011). However, broadly defined non-cognitive skills refer also to various other characteristics (e.g. reliability, capacity to cooperate, resilience and standards of professional performance) which command a direct return in the labor market. The type of heterogeneity in crime preferences that we model captures only a subset (with no direct market returns) of the broader non cognitive heterogeneity. For this reason we are hesitant to interpret these crime preference shocks as a proxy for the general non-cognitive skills.

in ability and education, (ii) persistent shocks to individual labor productivity, (iii) a deterministic trend in labor productivity over the life cycle, and (iv) asset accumulation. In the absence of the crime preference shock χ, the model would attribute all the variability in the data to economic factors. The introduction of an exogenous source of heterogeneity implies that both economic and non-economic factors can be the driver of crime choices in the model. The features of the data that help to distinguish the contribution of these two sets of factors, and identify the associated parameters, are discussed in Section 4.

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

3.3 Technology and markets

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

$$Q(H, K) = H^{1-\phi} K^{\phi}, \qquad 0 < \phi < 1, \tag{8}$$

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

$$H = [s_L H_L^{\varrho} + s_H H_H^{\varrho} + (1 - s_L - s_H) H_C^{\varrho}]^{\frac{1}{\varrho}}. \tag{9}$$

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 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 a_j^{nat} the natural borrowing limit.[24] There are perfect annuity markets to insure against mortality risk.

[24]The calibrated exogenous borrowing limit $\underline{a}$ is tighter than the natural borrowing limit 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.

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.

3.3.1 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, i.i.d. shock $\upsilon_j \in \{0, \kappa\}$ to labor income with $\pi_\upsilon = Pr\{\upsilon_j = \kappa\}$ 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.

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

$$\log h_j(\theta, e, \varepsilon_j^e) = \gamma^e \theta + \zeta_j^e + \varepsilon_j^e, \tag{10}$$

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

$$\varepsilon_j^e = \rho^e \varepsilon_{j-1}^e + \eta_j^e, \qquad \eta_j^e \overset{iid}{\sim} N(0, \sigma^e). \tag{11}$$

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

$$a_{j+1} = a_j(1+r)\lambda_j^{-1} + \tau_j \kappa \overline{wh} + (1 - i_j^p)[(1 - \upsilon_j)\, w^e h_j \left(\theta, e, \varepsilon_j^e\right) - c_j], \tag{12}$$

where a_j is the stock of financial wealth, r the (post-tax) risk-free interest and i_j^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 an adjustment factor for the actuarially-fair annuity premium.

The first two addenda on the right hand side of equation (12) 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—$i_j^p = 0$—earn labor income $w^e h_j \left(\theta, e, \varepsilon_j^e\right)$, are robbed of a share υ_j of it and choose consumption c_j.

3.4 Government and education cost

Government: The government administers a pay-as-you-go pension system, the criminal justice system, spends on transfers and wasteful public expenditure, 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 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 tuition fee $tuit^e$ minus the government subsidy sub^e for each year attended.

3.5 The individual problem

We write the individual problem in recursive form. Let $\mathbb{E}$ denote the conditional expectation over the joint distribution of all next period's shocks and $\mathbb{E}_i$ the conditional expectation over the marginal distribution of the shock i. 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 $V_{e,j}^s(a_j)$ denote the value of being in education for an individual of age j, (omitted) type (θ, χ), completed education e and financial wealth a_j. Similarly,

$V^n_{e,j}(a_j, \varepsilon^e_j)$ denotes the value of working for an individual of age j, (omitted) type (θ, χ), completed education e, financial wealth a_j and labor efficiency shock ε^e_j.

The value function of a newborn with initial wealth a_0 satisfies

$$V^0_L(a_0) = \max\{V^s_{L,0}(a_0), \mathbb{E}V^n_{L,0}(a_0, \varepsilon^L_0)\}, \tag{13}$$

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

$$V^s_{e,j}(a_j) = \max_{c_j, a_{j+1}} u(c_j) + \psi^{e'}(\theta) + \xi_j V^s_{e,j+1}(a_{j+1}) \tag{14}$$

$$\text{s.t. } a_{j+1} = a_j(1+r)\lambda_j^{-1} - tuit^{e'} + sub^{e'} - c_j, \quad a_{j+1} \geq \underline{a}. \tag{15}$$

These students do not face a study-work choice at age $j+1$ and their continuation value is $V^s_{e,j+1}$. Among them, college students aged $j_C - 1$ will enter the labor market the following period and their continuation value satisfies the condition $V^s_{C,j_C}(a_{j_C}) = \mathbb{E}V^n_{C,j_C}(a_{j_C}, \varepsilon^C_{j_C})$.

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

$$V^s_{e,j}(a_j) = \max_{c_j, a_{j+1}} u(c_j) + \psi^{e'}(\theta) + \xi_j \max\{V^s_{e',j+1}(a_{j+1}), \mathbb{E}V^n_{e',j+1}(a_{j+1}, \varepsilon^{e'}_{j+1})\} \tag{16}$$
$$\text{s.t.} \quad (15).$$

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. After observing her current labor efficiency shock, the value function of a worker aged $j \neq j_b$, is

$$\begin{aligned} V^n_{e,j}\left(a_j, \varepsilon^e_j\right) = &\max_{\tau_j} \ \chi \mathbb{I}_{\tau_j} + \tau_j \pi_p V^p_{e,j}(a_j, \varepsilon^e_j) + \\ &(1 - \tau_j \pi_p)\mathbb{E}_v[\max_{c_j, a_{j+1}} u(c_j) + \xi_j \mathbb{E}V^n_{e,j+1}(a_{j+1}, \varepsilon^e_{j+1})] \\ &\text{s.t. } (12), \quad a_{j+1} \geq \underline{a}. \end{aligned} \tag{17}$$

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 entering jail at age j is

$$V_{e,j}^p\left(a_j, \varepsilon_j^e\right) = u(\bar{c})\left(1 + \sum_{i=j}^{j+P-1} \Xi_i\right) + \Xi_{j+P}\Big[f[u(\bar{c}) + \xi_{j+P}\mathbb{E}V_{e,j+P+1}^n\left(a_{j+P+1}, \varepsilon_{j+P+1}^e\right)]+ (1-f)\mathbb{E}V_{e,j+P}^n\left(a_{j+P},, \varepsilon_{j+P}^e\right)\Big], \tag{18}$$

where $\Xi_i = \prod_{s=j}^{i-1} \xi_s$ is the discount factor from the current age j to age i.[25] $P \geq 1$ and $0 < f < 1$ denote, respectively, the integer and the fractional part of the prison term, in years. The value of entering jail is the present value of the flow of utility $u(\bar{c})$ over the integer part of the prison term plus the continuation value associated with the fractional part of the prison term. We model the fractional part of the prison term as a lottery which assigns probability f to the value of spending an extra year in prison and the complementary probability to the value of exiting at the end of the period, as in Cozzi (2006).[26] The reason for modeling fractional prison terms this way is that reducing the period length to an appropriately small unit (e.g. one or two months) would increase 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.

[25] Note the convention that $\Xi_j = 0$.

[26] 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.

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

$$
\begin{aligned}
V^n_{e,j}\left(a_j, \varepsilon^e_j\right) = & \max_{\tau_j} \ \chi \mathbb{I}_{\tau_j} + \tau_j \pi_p V^p_{e,j}(a_j, \varepsilon^e_j) + \\
& (1 - \tau_j \pi_p) \mathbb{E}_v [\max_{c_j, a_{j+1}, b} u(c_j) + v(b) + \xi_j \mathbb{V}^n_{e,j+1}(a_{j+1}, \varepsilon^e_{j+1})] \qquad (19) \\
\text{s.t. } & a_{j+1} = a_j(1+r)\lambda_j^{-1} + \tau_j \kappa \bar{w} + (1 - i^p_j)[(1 - \upsilon_j) w^e h_j \left(\theta, e, \varepsilon^e_j\right) - c_j - b], \\
& a_{j+1} \geq \underline{a}, \quad b \geq 0.
\end{aligned}
$$

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 given by equation (18) adjusted for the utility from (the zero) bequest $v(0)$ appropriately discounted.

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

$$
\begin{aligned}
V^n_{e,j}(a_j) = & \max_{c_j, a_{j+1}} u(c_j) + \xi_j V^n_{e,j+1}(a_{j+1}), \qquad (20) \\
\text{s.t. } & a_{j+1} = a_j(1+r)\lambda^{-1} + pen - c_j, \quad a_{j+1} \geq \max\{\underline{a}, \underline{a}^{nat}_j\}.
\end{aligned}
$$

4 Parameterization

We now turn to the description of the model's 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, together with the set of targeted moments, is listed in Table 2. 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 numerical model are expressed as a share of average labor earnings in 1980.[27] We report them in 1980 dollars in what follows.

Demographics. Each period represents one year. The real world counterpart of the first

[27]This amounted to 12,400 dollars.

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 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 (9) are based on estimates by Abbott, Gallipoli, Meghir, and Violante (2013) using CPS and PSID data. For ϱ, we use their preferred estimate of $\varrho = 0.68$, corresponding to an elasticity of substitution of 3.1.[28] The share parameters are calibrated using an average of the values estimated for the period 1979-1981 and are reported in 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 *pen* is assumed to be a constant lump sum for all agents, regardless of their education and previous earnings. The pension is set to 1,980$ 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 9,240$ per year. This corresponds to an average (across jails and state prisons) yearly cost of incarceration of 8,300$ per prisoner in 1980[29] plus an annualized total court-and-police cost per-convicted property criminal of 940$.[30]

[28]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.

[29]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.

[30]According to Table IV-D in Aos, Phipps, Barnoski, and Lieb (2001) court-and-police costs equal approximately 18% of incarceration costs per-prisoner. Since the average prison length is 19 months, this corresponds to 11.4% (18% $\times$ 19/12) of prison costs on an annualized basis. Aos, Phipps, Barnoski, and Lieb (2001) is the only study that provides a detailed breakdown of *all* the costs associated with the most common types of crime. The only limitation is that their estimates are based on data from the Criminal Justice System in Washington rather than all U.S. states.

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 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 (10) that the model implies the specification

$$\log W^e_{ijt} = \log w^e + \gamma^e \theta_i + \zeta^e(j_{it}) + \varepsilon^e_{ijt}. \tag{21}$$

for the legal earnings of an individual i, with education e, at age j and time t

The values of all the parameters in equation (21) are taken from the estimates in Abbott, Gallipoli, Meghir, and Violante (2013). The measure of ability θ_j, both in the model and in the data, is the deviation of the log of the AFQT89 test score from its mean and is based on the NLSY79.[31] The estimates for the ability gradient are 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.

[31] 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.

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

$$\varepsilon^e_{ijt} = \rho^e \varepsilon^e_{i,j-1,t-1} + \eta^e_{ijt}, \qquad \eta^e_{ijt} \overset{iid}{\sim} N(0, \sigma^e) \tag{22}$$

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 pure time discount factor $\hat{\xi}$ 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)[32] The implied value is $\hat{\xi} = 0.967$.

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

$$v(b) = \nu_1 \left(\nu_2 b\right)^{\nu_3} \tag{23}$$

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 (2013).[33] 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.

[32] It is well known that the PSID and NLSY, from which most of our other estimates are obtained, significantly 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.

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

Borrowing Limit. The exogenous borrowing limit $\underline{a}$ 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 U.S. economy, as measured by the CPS for 1980.[34] For aggregate consistency we adjust the NLSY79 rates so that they aggregate to the average CPS education rates observed between 1977 and 1983.[35] 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.[36]

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),

[34] 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.

[35] 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 (2013) to estimate the wage equations and the production technology.

[36] 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.

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[37] 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.[38]

We set the exogenous consumption level in prison $\bar{c}$ to match a property victimization rate of 5.6 per cent in 1980.[39] The resulting value is $\bar{c} = 3,900$ dollars.

Individual fixed effects. For computational simplicity the distribution of the innate ability fixed effect θ 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—a rather flexible parametric distribution—with support $(\underline{\chi}, 1+\underline{\chi})$ and exponents $\{\alpha, \beta\}$.

We allow for an arbitrary correlation $\rho_{\theta\chi}$ between the ability and crime fixed effects. We do so through a normal copula.[40] 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 crime participation regression on both simulated data and on 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.[41] The measure of property crime we use is whether or not the respondent reported having engaged in shoplifting, or stealing something worth 50$ or more, from

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

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

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

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

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

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%.[42]

The four model parameters determine the relative contribution of the economic and non-economic crime motive in the model. 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.

4.1 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. In what follows we provide a discussion of the relation between subsets of moments and parameters. The fact that some parameters have little or no influence on specific sets of model moments makes identification effectively block recursive.

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, in each iteration of the search for the equilibrium of the benchmark economy, to match the education distribution within the five ability bins.

Wealth moments. A second set of parameters—consisting of the pure time discount factor $\hat{\xi}$, 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 affect 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

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

ν_3 changes both the variance of bequests and their average, while ν_1 mostly influences the average. The discount factor $\hat{\xi}$ 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 crime moments by changing the wealth distribution, the reverse is not true. The remaining parameters are the share of income lost to crime κ, consumption in jail $\bar{c}$, the correlation $\rho_{\theta\chi}$ between ability and crime preference shocks and the parameters $\{\alpha, \beta, \underline{\chi}\}$ of the Beta distribution of the crime preference shocks. These parameters influence primarily the seven crime moments and are (over-) identified by them for a given value of the other parameters. Since κ is directly determined by the average loss from property crime, we are left with the five parameters $\{\bar{c}, \rho_{\theta\chi}, \alpha, \beta, \underline{\chi}\}$ that are identified by the last six moments in Table 2. Of these six moments the last four are estimated from a crime regression in which the dependent variable is a crime participation dummy; therefore these moments are related to the extensive—participation—margin of the crime choice. By contrast, the first two moments in the table—namely, the victimization rate and the share of high school dropouts among inmates—are also affected by the intensive margin of the crime choice, i.e. the number of crimes conditional on engaging in crime. The distinction between intensive and extensive margin of crime is important for identification: the intensive choice responds to the marginal net benefit of one extra crime which depends, positively, on the income flow from each crime and, negatively, on the utility loss associated with the increased probability of apprehension due to committing more crimes. In turn, the utility loss due to apprehension decreases with the flow of consumption in jail $\bar{c}$ and increases with legal market returns. The extensive participation margin is different because it responds to the total net benefit which also includes the *fixed* utility cost χ. For these reasons the extensive margin of the crime choice changes with both economic and non-economic incentives, while the intensive margin (conditional on participating in crime) is a function of economic incentives only. In order to clarify how joint identification is achieved it is helpful to consider the parameters sequentially. Suppose, for the time being, that the correlation $\rho_{\theta\chi}$ between ability and the crime taste shock is zero. The parameters $\{\alpha, \beta, \underline{\chi}\}$ and $\bar{c}$ jointly determine the relative importance of the economic versus the non-economic crime motive. If the distribution of χ were degenerate at zero, all crime would be economic and the flow of consumption in jail $\bar{c}$ would be identified by the aggregate victimization rate. However in such a world the share of high school dropouts among inmates would be much higher (by 20 percentage points) than in the data. The education and age coefficients would also be much larger (respectively one-half and five times larger, in absolute value) than in the data, while the ability coefficient would be *negative*. This establishes the need for some exogenous, non-economic heterogeneity in crime participation.

Under the assumption that the heterogeneity in the crime taste χ has a Beta distribution, the three parameters $\{\underline{\chi}, \alpha, \beta\}$ uniquely determine the first three moments of such distribution. The shape parameters α and β jointly determine the variance and skewness of the Beta distribution and govern the amount of heterogeneity in the relative preference for crime.[43] Given $\{\alpha, \beta\}$, the location parameter $\underline{\chi}$ uniquely determines the mean of the distribution. This allows us to discuss identification in terms of the first three moments of the crime preference shock distribution, which have a clearer economic intuition.

The first three moments of the crime taste distribution are over-identified by some data moments which only depend on the extensive margin of the crime choice (the intercept, education and age slopes in the crime participation regression) and by the share of dropouts among inmates which also reflects the intensive margin. In particular, the variance and skewness of the crime preference shocks induce heterogeneity in crime participation among individuals with similar economic circumstances. For instance, if the variance were low then crime participation differences would be mostly driven by economic differences, and non-criminals would on average have considerably higher wages than criminals. As preference shocks become more dispersed the amount of non-economic heterogeneity increases and, as a result, the difference in crime participation between better-off and poorer individuals is attenuated. In fact the variance of the non-economic motive χ has to be large enough to reduce the age gradient in the crime participation regression from a counter-factually high value of -3.5, in the absence of heterogeneity, down to a target value of -0.71 in the data. The relatively large variance implied by our

[43] It can be shown that, regardless of location $\underline{\chi}$, there is a one-to-one relationship between the second and third moment of the Beta distribution and each $\{\alpha, \beta\}$ pair.

estimate suggests that there are both low wage individuals who do not engage in crime and some high wage individuals who do participate in crime.

On the other hand, the skewness relocates probability mass to either side of the mean of the distribution. An increase in skewness increases the proportion of individuals with below-average crime taste and, therefore, reduces the number of infra-marginal criminals while increasing that of infra-marginal non-criminals. For this reason the third moment of the crime shock distribution is identified by the intercept of the calibration regression which corresponds to the average crime participation rate when all regressors are set to zero. The value of the intercept conveys information about that fraction of total crime participants who engage in criminal activity irrespective of changing economic incentives.

Turning to the location parameter $\underline{\chi}$, consider changing it while adjusting the value of $\bar{c}$ to keep the victimization rate constant. This alters the breakdown of the total cost of crime between the average (fixed) utility cost of crime and the variable opportunity cost associated to punishment. It follows that for given coefficients in the crime participation regression, which reflects only the extensive participation margin, $\underline{\chi}$ is identified by the share of dropouts among inmates—which responds to both the extensive and intensive margin of the crime choice. In other words, there is only one combination of $\underline{\chi}$ and $\bar{c}$ which is consistent with the observed victimization rate while also matching the share of dropouts among inmates.

Finally, the correlation $\rho_{\theta\chi}$ between ability and the crime fixed effect is identified by the coefficient of ability and, to a lesser extent, education in the crime participation regression used for the calibration. To see this suppose that all crime were economic—the crime fixed effect is zero for all agents—and $\rho_{\theta\chi}$ were, therefore, trivially zero.. As discussed above, the ability gradient and the coefficient of the education dummy in the regression would be negative because they both reduce the net benefits from crime. Consider now the opposite extreme case, in which all crime is non-economic and driven by the randomly distributed taste shock χ. If $\rho_{\theta\chi}$ were zero the ability gradient in the regression would be zero. If instead $\rho_{\theta\chi}$ were different from zero, then the ability gradient would have the same sign as $\rho_{\theta\chi}$: for example, if $\rho_{\theta\chi}$ were positive it would mean that more able people would on average have higher crime participation. Therefore, in order to match the positive ability coefficient in the calibration regression the model needs a positive correlation $\rho_{\theta\chi}$. The fact that the ability and education coefficients have opposite signs is further indication of the concurrence of both economic and non-economic incentives in criminal behavior.

5 Numerical Analysis

This section discusses the benchmark equilibrium and presents the results of our policy experiments. Section 5.1 overviews the implications of the model for a range of non-targeted data moments. Section 5.2 describes the policy experiments. In all the experiments, the proportional tax rate on labor income adjusts to balance the government budget. Section 5.3 conducts a sensitivity analysis with respect to the elasticity of input substitution among different education groups and the size of the borrowing constraint. Further 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 noting 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 response of high school and college enrollment to subsidies, the impact of education on incarceration rates and the age profile of arrest rates. We also evaluate the joint response of education and crime rates to various changes occurring between 1980 and 2000; we consider changes in the 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.

Enrollment responses. The model response of high school and college enrollment to conditional cash transfers appears to be consistent with existing evidence from available studies. The studies we consider concern either small-scale quasi-experimental interventions or large-scale simulated interventions in partial equilibrium. Therefore 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.

Dearden, Emmerson, Frayne, and Meghir (2009) measure the effect of subsidizing attendance to (post-compulsory) high school: they use data from a UK pilot study (the Education Maintenance Allowance) which offered means-tested conditional cash transfers for 16-to-18 year old to stay in full-time education.[44] 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.[45] A similar subsidy in the model increases high school attendance by a remarkably similar 6.5 percentage points.

In addition, we also compare the model's high school enrolment response to an unconditional high school graduation bonus of the same size—25,000$ (in 1994 dollars)—considered in the structural, partial equilibrium analysis of Keane and Wolpin (2000), which uses the NLSY79. This subsidy reduces the share of high school dropouts in the model by 22 percentage points, which is close to the corresponding response of 23 percentage points estimated by Keane and Wolpin (2000).[46]

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

[44]To the best of our knowledge this is the only empirical study of a randomized experiment to subsidize high school completion in a developed country. Attanasio, Fitzsimmons, Gomez, Gutiérrez, Meghir, and Mesnard (2006) and Attanasio, Meghir, and Santiago (2012) study the effects of randomized cash subsidy programs respectively in rural Colombia and rural Mexico. They find substantial effects of the policy. In another related paper Angrist and Lavy (2009) study an experiment subsidizing completion of the Israeli *Bagrut*. This experiment focuses on a fairly selected group because the Bagrut is a very academic secondary degree and a formal prerequisite for college admission. The estimated average effect in the sample is also positive.

[45]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.

[46]This is the fall in the share of dropouts over the whole population associated with a fall of 22 percentages points for whites and 29.7 for blacks reported in Keane and Wolpin (2000)

enrollment response to a 1,000$ (in 2001 dollars) change in college tuition costs.[47] 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 3.3 percentage points in response to a reduction in tuitions and a fall of 4.2 points in response to an increase, both within the range of existing estimates.

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 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.[48] 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), corresponding to a reduction of approximately half a percentage point in the probability of incarceration.[49] Finally, running the same regression on model-generated data and using the same set of common controls,[50] 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), corresponding to a reduction of 5.6 percentage points in the probability of crime participation. Estimating

[47] 1,000$ in 2001 correspond to 420$ in 1980.

[48] Lochner and Moretti (2004) 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."

[49] When we estimate the same regression also for white men alone, as in Lochner and Moretti (2004), the estimated coefficient is -0.39 (standard error 0.36).

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

the same regression using simulated data, with the same set of common controls plus the crime fixed effect,[51] 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.[52]

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 U.S. data.[53] The data are available for each individual age until age 24 and for five-year age brackets (e.g. 25-29) from age 25 onward. Arrest rates decline with age in both the data and model, reflecting the increasing opportunity cost of crime due to the upward sloping age profile of earnings and to life-cycle wealth accumulation. The model provides quite a good match until middle-age, but overstates arrest rates after age 45.

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:

- the high school and college shares changed respectively from 0.41 to 0.39 and from 0.37 to 0.45;[54]

[51]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.

[52]This is consistent with Lochner and Moretti's (2004) findings that "... a significant part of the measured effect of education on crime can be attributed to the increase in wages associated with schooling."

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

[54]Estimates by Abbott, Gallipoli, Meghir, and Violante (2013) available from the authors

- 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;[55]
- 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 theaverage 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 27.4 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).[56] *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 2.2%. The increase in the variance of income shocks and the change in demographics have effects which are similar in magnitude, but opposite in sign.

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

[55]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.

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

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 vs. tougher sentencing

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.[57]

In this section we use the model to compare the effects of two alternative policies—subsidizing high school completion and tougher sentencing—keeping constant their impact in terms of crime reduction. In each case, we assume the cost of the policy is financed by adjusting the labor tax rate and compare the policy implications in terms of financing costs, net output and welfare.

5.2.1 Subsidizing high school completion

In the experiments in this section we consider a yearly subsidy equal to 8.8 per cent of average labor earnings paid to *all* individuals attending and completing high school, independently of their financial means. 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.[58] 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.

Table 9 reports the changes in the education shares relative to the benchmark economy. The aggregate share of high school dropouts falls by one percentage point. The overall fall is unevenly distributed across ability bins. The share of high school dropouts actually *increases* in the lowest two ability bins, while it falls in the top three. The asymmetric response across ability bins reflects the increase in the relative price of high school

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

[58] 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.

dropouts and the complementarity between ability and education. For lower ability individuals, the fall in the return to schooling more than offsets the cost reduction due to the subsidy.[59] As a result, average ability falls among high school dropouts and increases among high school graduate. The latter effect is absent from the simple model in Section 2 which features no wealth heterogeneity, and reflects the fact that the subsidy reduces the extent of selection on the basis of wealth among the individuals with higher ability.

Column (1) in Table 10 reports a number of statistics for the benchmark equilibrium. Output and aggregate consumption (the net flow of consumable resources) and welfare in the benchmark equilibrium 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. Columns (3) reports the same statistics for the experiment under consideration.

The high school subsidy reduces the victimization rate from 5.6 to 5.2 per cent, an 8 per cent reduction. 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.[60] Since the apprehension probability is linear in the number of crimes 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 falls for both education groups, but the fall is particularly pronounced—more than 10 per cent—for high school dropouts. This fall among high school dropouts reflects the general equilibrium increase in the price of their labor that more than compensates the positive impact that would, otherwise, be implied by the fall in the average ability of their pool. Conversely, in the case of high school graduates the (minor) increase in the price of their labor and the improvement in their average ability both work in the direction of reducing their arrest rate. These findings are consistent with the mechanisms highlighted in the analytic model of Section 2: changes in the opportunity cost of crime appear to

[59] The smaller fall in the upper two bins relative to the third one is fully accounted for by their lower share of dropouts in the benchmark. In fact, such share drops to zero as a consequence of the subsidy.

[60] We do not report arrest rates for college graduates as their crime rate is negligible in the model.

have non-negligible effects on crime behavior, especially at the lower end of the earnings' distribution.

The improved education sorting by ability implies an increase in both output and efficiency, as measured by aggregate consumption, of roughly 1 percentage point. The increase in welfare, as measured by the ex ante consumption equivalent, is more than twice as large, amounting to a sizable 3.6 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 0.30 to 0.51 per cent of aggregate consumption in the benchmark. Despite this, the labor tax rate (not reported) marginally falls to 26.8 per cent due to the increase in the aggregate stock of human capital.

We also experiment with a subsidy equal to twice the original one. These results are reported in column (4) of Table 10. The increase in the subsidy reduces the aggregate share of high school dropouts by an additional 0.5 percentage points. Although 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 per cent relative to the benchmark. Note that while the total expenditure increases with the size of the subsidy, up to a sizable 0.72 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.

5.2.2 Tougher sentencing

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 analyzed in the previous section. 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 20.7 months achieve the same victimization rates as subsidies equal to, respectively, 8.8 and 17.6 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.

5.2.3 Subsidizing high school completion: partial equilibrium

In this subsection we study the effect of the same unconditional high school subsidy considered in Section 5.2.1 within a partial equilibrium setting, keeping all prices at their level in the benchmark economy. This exercise has two purposes. First, to gauge the quantitative importance of general equilibrium effects. Second, to provide an indirect test of the ability of the model to generate sensible aggregate labor income responses to the implied enrollment changes.

Unlike the partial equilibrium enrollment experiments in Section 5.1, we now study the steady-state equilibrium in which the distribution of physical and human capital is stationary. This provides the appropriate counterpart to the general equilibrium experiments in the previous sections, isolating the direct effect of the subsidy on the victimization rate from its indirect effect due to price changes.

Partial versus general equilibrium. Table 11 is the partial equilibrium counterpart of Table 9 and reports the change in the education shares relative to the benchmark, for a high school subsidy equal to 8.8 per cent of average labour earnings. The aggregate share of high school dropouts falls by 9 per cent as opposed to 1 per cent in general equilibrium. In the absence of wage adjustment, not only does the share of high school dropouts in the first two ability bins no longer rises, but its fall is larger than for the top three. As a result, average ability falls among both high school dropouts *and* high school graduates.[61] Turning to column (2) in Table 10, the fall in average ability among both

[61]A comparison with the benchmark education distribution in Table 7 makes clear that only people in the lowest two ability bins compose the pool of high school dropouts.

high school dropouts and high school graduates results in an approximate 10 per cent increase in the arrest rate within both groups, as opposed to the lower arrest rates prevailing in general equilibrium. As a consequence the aggregate victimization rate drops only down to 5.4 per cent, compared to 5.2 in column (3); this smaller fall in crime reflects only the composition effect due to people moving between education groups with different crime rates, rather than a lower crime rate within each group.[62]

The increase in welfare is still large—3.4 per cent—and comparable to its value in general equilibrium. Prison expenditure falls marginally less and total expenditure on prison and education subsidies increases marginally more than in general equilibrium; the latter fact follows from the substantially higher subsidy take-up, driving up total policy costs from 0.30 to 0.54 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 vast increase in the aggregate stock of human capital.

Labor income responses. In Section 5.1 we have shown that the model generates enrollment responses to high school subsidies which are consistent with empirical estimates; however direct evidence on whether the associated output and welfare responses are plausible is not available. In fact, the difficulty of estimating long-run equilibrium effects constitutes an important reason to develop a general equilibrium model.

Nonetheless, we may compare the partial equilibrium increase in aggregate labour income in the model with the change implied by empirical estimates of the return to high school graduation. For example, Lochner and Moretti (2004) estimate a private return to high school graduation of 8,040$ per year in 1990. This corresponds to roughly 4,700$ in 1980 or approximately one third of average earnings in the model. Given an increase in high school enrolment of 9 percentage points, the associated increase in aggregate labor income would be 3.4 percentage points. In contrast, the change in aggregate labor income generated by the model is a conservative 2.4 percentage points.[63]

[62] As discussed in Section 2, the change in the education composition, though, reduces the aggregate victimization rate, as high school graduates, for any given ability, have a higher opportunity cost of crime—hence a lower crime rate—relative to high school dropouts due to the higher price of their human capital.

[63] Extrapolating the effects of large policy changes on the basis of (local) reduced-form elasticities may be misleading. The instrumental variables' analysis identifies the causal effect of high school graduation on earnings and provides an estimate of the conditional average effect of education. Using these estimates to approximate the aggregate labor income change due to higher high school completion implicitly

5.3 Sensitivity: labor substitutability and borrowing constraint

The analytic section emphasizes how the effects of education subsidies on crime depend on the relative price and the degree of substitutability in production between different education types. Moreover, access to credit may also play a role in shaping the policy response.

In what follows we study the same policy considered is Section 5.2.1—a high school subsidy equal to 8.8 per cent of average earnings—for different combinations of the degree of labor substitution and of the borrowing limit.[64]

We experiment with values of ρ equal to 0.5 and 0.8, which imply an elasticity of substitution among education groups of respectively 2 and 5. 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. As a comparison the benchmark value of $\rho = 0.68$ in the previous two sections corresponds to an elasticity of 3.1. For the borrowing limit, we experiment with three levels corresponding to a half, the same and twice its benchmark value.

For each combination of the three values of ρ, $\{0.5, 0.68, 0.8\}$, and of the borrowing limit, Table 12 reports the levels of the victimization rate, prison expenditure and subsidy expenditure as well as changes relative to the respective benchmark as a result of the high school subsidy experiment in general equilibrium.[65] Further relaxing the borrowing limit does not substantially change the responses.

The victimization rate only changes very little with the borrowing limit and the elasticity of substitution. As one would expect, for given ρ the welfare gains from the policy fall as the borrowing constraint is relaxed. Moreover, the sensitivity of the welfare response to the borrowing limit is increasing in the elasticity of input substitution. As noted in Section 2, the high school premium and the enrollment response to the subsidy are increasing in the substitution elasticity.[66] Tightening the borrowing limit changes

assumes that the additional high school graduates have the same average ability as those in the absence of the policy. The model results, instead, draw attention to the fact that the individuals who complete high school under the policy, but not in its absence, are of lower average ability. This accounts for the smaller increase in aggregate labor income.

[64] As in Section 5.2.2 above, changes in the sentence length have very little effect on output and welfare. Results are available upon request.

[65] All the alternative benchmarks are recalibrated.

[66] The response of high school enrollment is about two percentage points larger when ρ changes from

the welfare gain from high school subsidization relatively more when the return to high school is higher.

Finally, the output and aggregate consumption response are not very different across parameterizations, nor are the responses of expenditures associated with the policy. The price response is, as expected, decreasing in the substitution elasticity.

6 Additional sensitivity analysis

In this Section we report some additional sensitivity experiments with respect to parameters not discussed in Section 5.3. In particular, we consider changes in human capital shares in production, the distribution of initial resources, education costs and the correlation of unobserved characteristics. For each calibration, we report the equilibrium victimization rate and welfare in Table 13.

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[67] and confirm the conjecture based on the analytic example: for a larger skill premium the same education subsidy generates a larger drop in the victimization rate: 5.0% rather than 5.2% 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 ν_1 and ν_2 of the warm glow utility from bequests, which shape the wealth distribution at the start of life. Specifically, we change ν_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 ν_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

0.5 to 0.8.

[67] 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 rate drops by more. As expected, the subsidy is more effective in a context in which people are more constrained by initial resources.

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 high school costs and setting them to zero, again with no noticeable change in results.

7 Conclusions

We use a dynamic equilibrium model to compare two alternative sets of policies targeting crime reduction: subsidies for high school completion and increases in prison sentences. We find that, for the same 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.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. L. Violante (2013): "Education Policy and Intergenerational Transfers in Equilibrium," NBER Working Paper 18782.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. D. Kautz (2011): "Personality Psychology and Economics," NBER Working Paper 16822.

Angrist, J., and V. Lavy (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99(4), 1384–1414.

Aos, S., P. Phipps, R. Barnoski, and R. Lieb (2001): "The Comparative Costs and Benefits of Programs To Reduce Crime," Discussion paper, Washington State Inst. for Public Policy, Olympia WA.

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

Attanasio, O., E. Fitzsimmons, A. Gomez, M. I. Gutiérrez, C. Meghir, and A. Mesnard (2006): "Child education and work choices in the presence of a conditional cash transfer programme in rural Colombia," IFS Working Papers W06/13, Institute for Fiscal Studies.

Attanasio, O. P., C. Meghir, and A. Santiago (2012): "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA," *The Review of Economic Studies*, 79, 37–66.

Barillas, F., and J. Fernández-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.

Blumstein, A., and H. J. Beck (1999): "Population Growth in U.S. Prisons, 1980–96," *Crime and Justice*, 26, 17–62.

Blundell, R., M. C. Dias, and C. Meghir (2003): "The Impact of Wage Subsidies: A General Equilibrium Approach," Mimeo, UCL.

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.

Buonanno, P., and L. Leonida (2006): "Education and crime: evidence from Italian regions," *Applied Economics Letters*, 13, 709–713.

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.

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

Dearden, L., C. Emmerson, C. Frayne, 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., and P. A. LaFontaine (2010): “The American High School Graduation Rate: Trends and Levels,” *Review of Economics and Statistics*, 92, 244–262.

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.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24, 411–482.

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.

Keane, M. P., and K. I. Wolpin (2000): "Eliminating Race Differences in School Attainement and Labor Market Success," *Journal of Labor Economics*, 18(4), 614–652.

Lee, D. (2005): "An Estimable Dynamic General Equilibrium Model Of Work, Schooling, And Occupational Choice," *International Economic Review*, 46, 1–34.

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.

Lise, J., S. Seitz, and J. Smith (2004): "Equilibrium Policy Experiments and the Evaluation of Social Programs," NBER Working Paper N. 10283.

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

——— (2011): "Chapter 2 - Nonproduction Benefits of Education: Crime, Health, and Good Citizenship," in *Handbook of The Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, vol. 4 of *Handbook of the Economics of Education*, pp. 183 – 282. Elsevier.

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., O. Marie, and S. Vujić (2011): "The Crime Reducing Effect of Education," *The Economic Journal*, 121, 463–484.

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

Mauer, M. (1999): "The Crisis of the Young African American Male and the Criminal Justice System," Discussion paper, Sentencing Project.

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 Analytic example: a simple two period model

Given the reservation rule (2), the expression for the second-period utility in equation (1) can be rewritten as

$$U_2^e(\theta) = \log(w^e\theta) + \max\{0, (1-\pi)[\varsigma - \varsigma_r^e(\theta)]\}.$$

Integrating over $[-\bar{\varsigma}, \bar{\varsigma}]$, and using the reservation rule, yields

$$\begin{aligned}\mathbb{E}U_2^e(\theta) &= \log(w^e\theta) + \frac{1-\pi}{2\bar{\varsigma}}\int\limits_{\varsigma_r^e(\theta)}^{a}[\varsigma - \varsigma_r^e(\theta)]d\varsigma = \log(w^e\theta) + \frac{1-\pi}{4\bar{\varsigma}}[\bar{\varsigma} - \varsigma_r^e(\theta)]^2 \simeq \\ &\simeq \log(w^e\theta) + \frac{1-\pi}{4\bar{\varsigma}}[\bar{\varsigma}^2 - 2\bar{\varsigma}\varsigma_r^e(\theta)] = \left(1-\frac{\pi}{2}\right)\log(w^e\theta) + \frac{(1-\pi)\bar{\varsigma}}{4} + \frac{\pi}{2}\log\bar{c},\end{aligned}$$

where the second line follows from the fact that the quadratic term in $\varsigma_r^e(\theta)$ is approximately zero and can be disregarded under the assumption that $\pi^2 \simeq 0$. The last equality follows from substituting $\varsigma_r^e(\theta)$ using the reservation rule.

Replacing for $\mathbb{E}U_2^e(\theta)$ in the expressions for $U_1^e(\theta)$, and setting $U_1^H(\theta) = U_1^L(\theta)$, one obtains equation (3). Differentiating (3) with respect to the subsidy yields,

$$\frac{d\theta^*}{dsub} = -\left[\frac{1}{\theta^*} + \left(1-\frac{\pi}{2}\right)\frac{d\log(w^H/w^L)}{d\theta^*}\right]^{-1} \tag{24}$$

which is equation (4) in the main text.

Given the production technology, and for $\underline{\theta}$ approaching zero, the skill premium can be written as

$$\log\left(\frac{w^H}{w^L}\right) = \log\left(\frac{1-s}{s}\right) + (1-\rho)\log\left(\frac{H_L}{H_H}\right) = \log\left(\frac{1-s}{s}\right) + (1-\rho)\log\left(\frac{\theta^{*2}}{1-\theta^{*2}}\right), \tag{25}$$

where the second equality follows from $H_L = \int_0^{\theta^*}\theta d\theta = \frac{\theta^{*2}}{2}$ and $H_H = \int_{\theta^*}^1 \theta d\theta = \frac{1-\theta^{*2}}{2}$. The above expression implies that the second addendum in equation (24) is positive and decreasing in ρ, from which Remark 1 follows.

To obtain the probability that a worker of type (e, θ) is a criminal we note that the reservation rule (2) implies

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

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

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

for $\underline{\theta}$ converging to zero.

Adding the two integrals above, and rearranging, one obtains the expression for the aggregate measure of criminals C,

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

which is equation (4) in the main text.

One can use the above expression and equation (25) to show that $\frac{\partial C}{\partial\theta^*}$ is the sum of three distinct terms, as reported below:

$$\frac{\partial C}{\partial\theta^*} = \frac{\pi}{2\bar{\varsigma}(1-\pi)}\left\{\left[\log\left(\frac{1-s}{s}\right) + (1-\varrho)\log\left(\frac{\theta^{*2}}{1-\theta^{*2}}\right)\right] + \frac{2\,(1-\varrho)}{1-\theta^{*2}} - \frac{\partial\log\left(w^H\right)}{\partial\theta^*}\right\}. \tag{26}$$

The term in square bracket is the partial equilibrium effect and coincides with the skill premium. It is positive as long as the skill premium is positive. The second addendum captures the general equilibrium increase in the skill premium at given w^H; i.e., the fall in w^L due to the increase in the share of dropouts θ^*. This term is also unambiguously positive. The third addendum reflects the general equilibrium response of w^H and is ambiguous in sign. However we find that this term is consistently small in numerical simulations and does not affect the sign of the above expression.

The second part of Remark 2 in the main text follows from: (1) the observation that the term in square bracket in (26) is increasing in ϱ, in the empirically relevant case in which $H_H/H_L > 1$; (2) the fact that the second addendum in (26) is decreasing in ϱ.

Finally, as ϱ goes to one, the second and the third term in equation (26) both go to zero.

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 $\{pen, G, sub^e, t_l, t_k\}$, tuition fees $tuit^e$ 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 (14)-(16) for $x = s$, equation (17) for $x = n$ and $j < j_r, j \neq j_b$, equation (19) for $x = n$ and $j = j_b$, equation (20) for $j \geq j_r$;
   - the decision rule $a_{j+1}^p(z_j^x)$ satisfies equation (12) with $i^p = 1$ and the associate value function $V_j^p(z_j^p)$ solves equation (18);
   - the decision rule b solves equation (19);
   - the education decisions $\{i_L^s(z_0^s), i_H^s(z_{j_H}^s)\}$ solve equations (13) and (16);
   - the crime decision $\tau_j(z_j^n)$ solves equation (17).
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[68]

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

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[69] $PRIS = \sum_{j<j_r} \int_{Z_j^n} m \, \pi_p \tau(z_j^n)) d\mu_j^n$ 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 $\mu = \{\mu_0^s, ..., \mu_{\bar{j}}^s; \mu_0^n, ..., \mu_{\bar{j}}^n\}$ 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 the state space $\mathbb{Z} = \prod_{j,x} Z_j^x$, the Cartesian product of all Z_j^x. The mapping $Q(Z, \mu)$

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

[69] 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.

is the transition function associated with the individual decisions, the law of motion for the shocks $\{\chi, \theta, v, i^p, \varepsilon_j^e\}$ 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. We 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 optimization method 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.[70]

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.

[70] Barillas and Fernández-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.

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

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

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

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

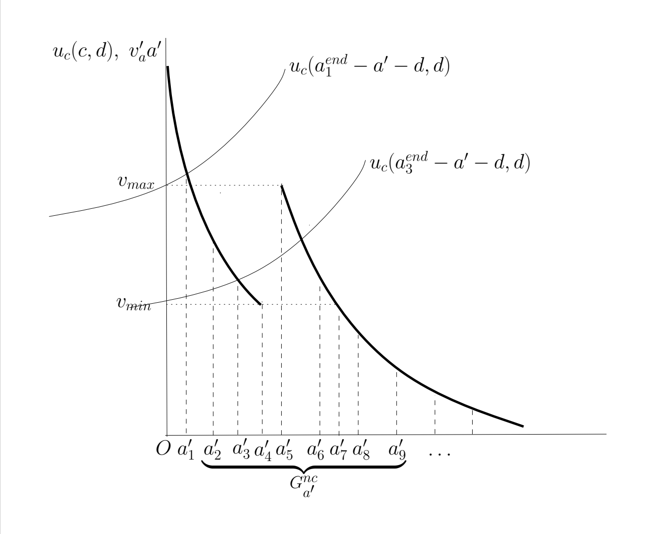
as $v'_a(a')$ can jump up but not down.[71]

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 $a'(a|d)$.

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

[71] 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



an ordered grid $G_{a'} = \{a'_1, a'_2, \ldots, a'_m\}$ for *end-of-period* assets a' and solve for the value of *initial wealth* a_i^{end} that satisfies EE for each $a'_i \in G_{a'}$. for each $a'_i \in G_{a'}$.[72] 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 $v'_a(a') \in (v_{min}, v_{max})$.[73] 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

1. Determine the non-concave region $G^{nc}_{a'}$. Initialize the counters $i = 1$ and $l = 1$
2. Solve EE for a_i^{end} given a'_i using EGM.

[72]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.

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

3. If $a'_i \in G^{nc}_{a'}$ then
   - find the maximizer of the discretized maximand for $a = a^{end}_i$; i.e. solve for

$$a'_g = \arg\max_{a' \in G^{nc}_{a'}} u(a^{end}_i - a' - d, d) + v'(a').$$

   - 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^{end}_i, a'_i) as $(a^{end}_{i_l}, a'_{i_l}) = (a^{end}_i, a'_i)$. 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 $\{a^{end}_{i_l}, a'_{i_l}\}$ on the endogenous collocation points $\{a^{end}_{i_l}\}$ solve for the conditional value function given d

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

6. Evaluate interpolating functions through (a^{end}_{il}, a'_{il}) and (a^{end}_{il}, 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[74]

[74]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 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.

for the same number of grid points.[75]

A.4 College subsidy

The main policy focus of our analysis has been the effect of high school subsidization. Such focus naturally relates to the existing literature—reviewed in Lochner (2011)—on the effects of schooling on crime. In this literature a motivation for early intervention is that the majority of property crime is committed by people with relatively low education, and high school graduation has been proven effective in reducing crime.

In this appendix, instead, we analyze the effects of a policy that subsidizes college completion. In particular, we consider a transfer paid to all people who enroll and complete college. For comparability with the other policy experiments in the main body of the paper, we consider a college subsidy that achieves the same general equilibrium crime reduction as the high school subsidy equal to 8.8% of average earnings studied in the paper. The size of the college subsidy that achieves the targeted victimization rate of 5.2% is about 15.3% of average earnings.

Table 14 reports the effects (relative to the benchmark) of the college subsidy in both partial and general equilibrium. In general equilibrium the policy costs as much as the high school subsidy policy, but it induces a somewhat smaller welfare gain. Intuitively a college subsidy provides less insurance against ex ante uncertainty, relative to the high school subsidy, as it primarily affects high ability individuals marginal to the college choice.

To understand what drives the general equilibrium response, it is instructive to consider what the effect would be in partial equilibrium. The policy actually *increases* the crime rate by almost 0.3 percentage points in partial equilibrium as it induces a very large rise in college completion. The size of the subsidy, together with the relatively low cost of college attendance in 1980, implies that college education becomes not only free at the point of entry, but is associated to a non-trivial monetary transfer. Such large shift in college completion, at constant prices, increases income inequality and the average return from crime. In contrast, the high school subsidy actually *reduces* the crime rate

[75] 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.

by a similar amount already in partial equilibrium. Therefore, all of the crime reduction effects of the college subsidy are due to general equilibrium effects. These are even larger than in the case of the high school subsidy, as the college subsidy substantially increases the human capital price not only for high school dropouts, but also for high school graduates. The resulting increase in earnings among the lowest two education groups raises the opportunity cost of engaging in crime for those agents who are most likely to commit crime. The difference between partial and general equilibrium is stark and highlights the importance of general equilibrium adjustments, which would be the only anti-crime justification for implementing such a policy.

Finally, one word of warning is necessary when assessing these results on the effects of a universal college subsidy. The direct costs of attending college have been steadily increasing over time, and continue to do so. As we mentioned above, a universal college subsidy of 15.3% (the one considered in this experiment) would have resulted, in 1980, in free college at the point of entry plus a yearly handout equal to roughly half the college cost. The same proportional subsidy in 2000, when college tuitions were more than double those in 1980, would instead have covered only between half and two thirds of the direct cost of college.

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	30	Bequest age (real world 45)
j_r	50	Maximum working life (real world 65)
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
$tuit^H$	124$	Direct cost of high school
$tuit^C$	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
$\psi^H(\theta), \psi^C(\theta)^a$		Enrollment 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: Abbott, Gallipoli, Meghir, and Violante (2013).

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: Abbott, Gallipoli, Meghir, and Violante (2013).

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 dollars). Source: Abbott, Gallipoli, Meghir, and Violante (2013).

	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: Abbott, Gallipoli, Meghir, and Violante (2013).

	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
L	86	25	11	3	1	25
H	13	70	78	73	54	57
C	1	5	11	24	45	17

Table 8: Model implications

Statistic	Data	Model
Enrollment response (pct. points)		
High school enrollment (Small subsidy)[a]	6.7	6.5
S.E.	(1.7)	
High school enrollment (Large subsidy)[b]	23	22
College enrollment[c]	3-9	3.3/4.2
Effect of education on incarceration[d]		
HS school graduation dummy	-0.49	-0.37
S.E.	(0.39)	
Effect of education on crime participation[e]		
HS school graduation dummy	-5.63	-4.07
S.E.	(2.25)	
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	27.4
Changing: Human capital shares (skill bias):		
Crime rate (%)	-	8.0
Share of high school graduates (%)	-	56
Share of college graduates (%)	-	32
Changing: Expected punishment:		
Crime rate (%)	-	2.2
Changing: Income variance:		
Crime rate (%)	-	5.8
Changing: College tuitions:		
Crime rate (%)	-	5.6
Share of high school graduates (%)	-	59
Share of college graduates (%)	-	15.4
Changing: Demographics:		
Crime rate (%)	-	5.3

Notes:

[a]The data numbers are taken from Dearden, Emmerson, Frayne, 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.

[b]The data numbers are taken from Keane and Wolpin (2000). Estimated response of high school enrollment to a high school subsidy equal to 25,000$ (1994 dollars).

[c]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.

[d]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.

[e]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.

Table 9: 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
L	4	1	-8	-3	-1	-1
H	-4	0	11	1	-5	1
C	0	-1	-3	1	6	0

Table 10: Subsidy and prison experiments. Subsidy as % of average labor 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	20.7
Crime victimization (%)	5.6	5.4	5.2	4.9	5.2	4.9
Arrest rate L (‱)	5.9	6.7	5.4	4.8	5.6	5.2
Arrest rate H (‱)	2.7	2.9	2.6	2.6	2.5	2.4
L share of criminals (%)	48	36	45	42	49	49
Output	100.0	-	101.1	101.9	99.9	99.9
Agg. Consumption	100.0	-	101.4	102.3	99.9	100.0
Welfare	100.0	103.4	103.6	106.6	100.0	100.0
Prison expenditure†	0.30	0.29	0.28	0.26	0.30	0.29
Subsidy + prison exp.†	0.30	0.54	0.51	0.72	0.30	0.29
Price L	100.0	-	102.8	104.9	100.0	100.0
Price H	100.0	-	100.4	100.9	100.0	100.0
Price C	100.0	-	99.7	99.5	100.0	100.0

† As a share of aggregate consumption in the benchmark.

Table 11: 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
L	-17	-14	-11	-3	-1	-9
H	17	14	9	-2	-7	6
C	0	0	2	5	8	3

Table 12: Results of sensitivity analysis for different values of the elasticity of input substitution and of the borrowing limit. Each column reports the levels of the victimization rate, prison expenditure and subsidy expenditure as well as changes relative to the respective benchmark as a result of the 8.8% high school subsidy experiment in G.E.

ρ (Labor elasticity)		0.5			0.68			0.8	
$\underline{a}$ (Borrowing limit)	-.49	-.99	-1.98	-.49	-.99	-1.98	-.49	-.99	-1.98
Crime victimization (%)	5.2	5.2	5.2	5.2	5.2	5.2	5.1	5.1	5.1
Output	101.0	100.8	100.7	101.1	101.1	100.9	101.6	101.1	101.1
Agg. Consumption	101.2	101.0	100.8	101.4	101.4	101.1	102.1	101.4	101.3
Welfare	103.8	103.4	102.7	104.8	103.6	102.6	104.8	103.2	102.6
Prison expenditure†	0.29	0.28	0.28	0.28	0.28	0.28	0.27	0.28	0.27
Subsidy + prison exp.†	0.52	0.51	0.51	0.50	0.51	0.50	0.51	0.51	0.50
Price L	103.0	102.7	102.5	102.9	102.8	102.4	102.7	102.3	101.9
Price H	100.3	100.3	100.4	100.4	100.4	100.4	100.5	100.3	100.4
Price C	99.7	99.7	99.8	99.7	99.7	99.8	99.7	99.6	99.9

† As a share of aggregate consumption in the benchmark.

Table 13: Results of sensitivity analysis. Each line corresponds to an alternative parameterization 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. 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	Experiment	
Labor shares - year 2000	5.6	5.0	103.5
Higher average inter-vivos	5.6	5.3	103.2
Lower average inter-vivos	5.6	5.1	104.5
Higher % zero initial wlth	5.6	5.1	104.3
Lower % zero initial wlth	5.6	5.2	102.8
Higher cost of High School	5.6	5.2	103.4
Lower cost of High School	5.6	5.2	103.2
Higher cost of College	5.6	5.2	103.4
Lower cost of College	5.6	5.2	103.1
Higher $\rho_{\theta\chi}$	5.6	5.2	103.6
Lower $\rho_{\theta\chi}$	5.6	5.2	103.7

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

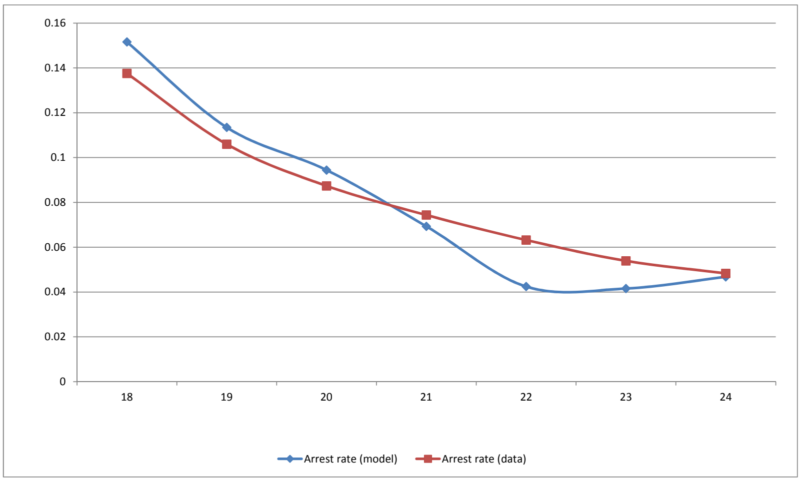


Figure 2: Five-year arrest rates for ages 20-55 in data and model.

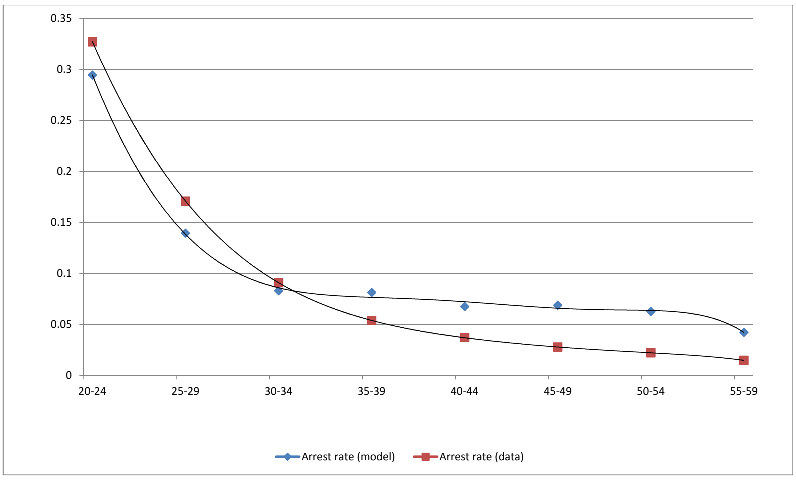


Table 14: College subsidy experiments. Subsidy as % of average labor income.

	Benchmark	COL Subsidy PE	COL Subsidy GE
	(1)	(2)	(3)
COL subsidy	-	15.3	15.3
Prison sentence (months)	19	19	19
Crime victimization (%)	5.6	5.9	5.2
Arrest rate L (‱)	5.9	7.1	5.5
Arrest rate H (‱)	2.7	3.4	2.7
L share of criminals (%)	48	51.5	47
Output	100.0	-	101.6
Agg. Consumption	100.0	-	102.1
Welfare	100.0	106.3	103.1
Prison expenditure†	0.30	0.31	0.28
Subsidy + prison exp.†	0.30	0.88	0.51
Price L	100.0	-	102.4
Price H	100.0	-	102.3
Price C	100.0	-	97.9

† As a share of aggregate consumption in the benchmark.