Education and Crime over the Life Cycle

Giovanni Gallipoli*

joint with

Giulio Fella†

January 2006

Job Market Paper

Abstract

In this paper we ask whether policies targeting a reduction in crime rates through changes in education outcomes can be considered an effective and cost-viable alternative to interventions based on harsher punishment alone. In particular we study the effect of subsidizing high school completion. Most econometric studies of the impact of crime policies ignore equilibrium effects and are often reduced-form. This paper provides a framework within which to study the equilibrium impact of alternative policies. We develop an overlapping generation, life-cycle model with endogenous education and crime choices. Education and crime depend on different dimensions of heterogeneity, which takes the form of differences in innate ability and wealth at birth as well as employment shocks. PSID, NIPA and CPS data are used to estimate the parameters of a production function with different types of human capital and to approximate a distribution of permanent heterogeneity. These estimates are used to pin down some of the model's parameters. The model is calibrated to match education enrolments, aggregate (property) crime rate and some features of the wealth distribution. In our numerical experiments we find that policies targeting crime reduction through increases in high school graduation rates are more cost-effective than simple incapacitation policies. Furthermore, the cost-effectiveness of high school subsidies increases significantly if they are targeted at the wealth poor. We also find that financial incentives to high school graduation have radically different implications in general and partial equilibrium (i.e. the scale of the programmes can substantially change its outcomes).

*University College London and IFS/CEMMAP. *Correspondence:* Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE. *E-mail:* g.gallipoli@ucl.ac.uk .

†Queen Mary, University of London. E-mail: g.fella@qmul.ac.uk.

We thank Sami Berlinski,Marco Cozzi,Lance Lochner,Costas Meghir,Jean Marc Robin and Nicola Pavoni for comments and discussions.

1 Introduction

Crime is a hot issue on the US policy agenda. Despite its significant fall in the Nineties its cost to the taxpayer has soared. The prison population has doubled over the same period and now stands at over two millions of inmates. The yearly cost of keeping a person in jail exceeds 20,000 dollars[3]. These numbers beg the question of whether the country is using a cost-effective mix of policies in the fight against crime.

For reasons similar to those highlighted in Heckman, Lochner, and Taber (1998), the analysis of alternative policies to tackle crime benefits from the use of a dynamic equilibrium framework[4]. Dynamics are important as educational and criminal choices interact in a life-cycle perspective. Furthermore, in equilibrium any large-scale programme is likely to alter personal disposable income both through its impact on market prices and through its financing implications. Therefore, in evaluating alternative programmes it is important to account for their impact on the personal distribution of disposable income and welfare.

This paper develops an empirically-based, heterogeneous-agent, equilibrium life-cycle model incorporating both education and criminal choices. Its goal is to provide a framework in which to analyse the effectiveness in terms of cost to the taxpayer and the welfare implications of alternative policies which directly or indirectly impact on crime. We apply the model to the study of property crime which is more likely to be driven by economic decisions, then, for example, homicide or rape.

The model has two main sources of heterogeneity. Agents differ in: 1) innate, observed ability and 2) initial wealth. Agents self-select themselves into education on the basis of these differences and, upon entering the labour market, decide whether to engage in criminal activity on a period-by-period basis.

We use PSID, NIPA and CPS data to estimate the parameters of a production function with different types of human capital and to approximate a distribution of permanent heterogeneity. These estimates are used to pin down some of the model's parameters. We also use PSID data to estimate the relative importance of ability in different education groups. The model is calibrated to match education enrolments both in the aggregate and by measured ability, aggregate (property) crime rates and some features of the wealth distribution.

[3]The figure comes from Maguire and Pastore (1995) and is at the lower end of available estimates. Section 4 reports alternative estimates.

[4]In fact, the case for the use of models allowing for equilibrium effects in policy analysis has been recently argued by various authors in different fields. See, among others, Abraham (2001), Lee (2001),Lee and Wolpin (2004), Cunha, Heckman, and Navarro (2004) and Gallipoli, Meghir, and Violante (2004)

We compare the implications and effectiveness of two policies: the first, an increase in the prison term by 1.2 months, the second, a subsidy towards high school completion of roughly 8% of average earnings per year of school. This amount is the same as the size of a well-known small scale program - the Quantum Opportunity Program - which provided extra support and high school graduation incentives aimed at children from a disadvantaged background[5]. We compare the effect of an unconditional subsidy paid to all high school graduates to a means-tested one aimed at students in the lowest 35% percentile of the wealth distribution. To control for heterogeneity we also experiment with assignment of agents to a treatment and control group.

The increase in the prison term reduces the aggregate crime rate by 4.2% and marginally expands the stock of inmates and the associated expenditure. As a consequence, the proportional labor tax rate has to increase marginally (from 27% to 27.03%) to finance the increased cost. The impact of the policy is effectively the same both in partial and in general equilibrium.

Financing an unconditional subsidy to high school completion calls for the same increase in the labor tax rate despite the fact that, in general equilibrium, the absolute cost of the intervention is twice as large as in the case of an increase in the prison sentence. The increase in efficiency and revenues makes the tuition policy basically self-financing. Furthermore, the tuition subsidy is more than twice as effective in terms of crime reduction. The associated fall in the crime rate is a sizeable 9%. The intuition behind the result is that, in equilibrium, the subsidy shifts lower ability people, who have a higher propensity to committing crime, out of the high school dropouts group. At the same time, it increases the proportion of people with relatively higher ability in the high school dropouts group. These are the people whose opportunity cost of attending high school goes up more in response to the increase in the relative wage of high school dropouts. Since education and skills are substitutes in their effect on crime rates, this reallocation is highly effective. The importance of this composition effect is apparent in partial equilibrium. While the policy induces a much larger fall in the number of high school dropouts, the ability composition of the pool of high school dropouts does not improve at unchanged prices. As a result the crime rate falls by only 3%.

The same subsidy paid only to students in the lowest 35% percentile of the wealth distribution goes a long way in reducing the crime rate at roughly one third of the cost relative to an equivalent unconditional subsidy. The aggregate crime rate falls by 6% in general equilibrium while the increase

[5]See Hahn, Leavitt, and Aaron (1994) and Taggart (1995) for a discussion of the program and its effects

in efficiency implies the policy can be financed at a marginally lower labor tax rate relative to the benchmark. The mechanism at play and the differences between partial and general equilibrium are similar to the unconditional subsidy experiment.

Conducting the same means tested experiment but randomized over a treatment and a control group allows us to compare the predictions of our model to the outcome of actual randomized programs such as the Quantum Opportunities Program. The average crime rate over the life cycle is between 14 and 15% for the control group and half as much for people in the treated group which took up the subsidy. This is broadly consistent with the findings by Hahn, Leavitt, and Aaron (1994) in their follow-up study among QOP participants one year after the end of the program. The proportion of people reporting being involved with the police at least once is 6% for the treated group against 13% among controls.

The model is in the tradition of economic models of crime which goes back to Becker's (1968) seminal contributions. There is an extensive body of empirical literature testing the main prediction of the theory that both market returns and the expected punishment are significant determinants of criminal choices. The effect of market returns upon crime is well established: see for example Grogger (1998) and Freeman (1999) who surveys earlier empirical studies. More recent work includes Gould, Weinberg, and Mustard (2002), Machin and Meghir (2004) and Raphael and Ludwig (2003). Concerning the effect of the expected punishment, Levitt (1997) and Levitt (1996) finds significant elasticity of crime rates respectively to expenditure on police and the length of the expected prison term. Finally, the existence of a relationship between crime and education is documented by Lochner and Moretti (2004).

Imrohoroglu, Merlo, and Rupert (2004) are the first to study jointly the effect of changes in market returns and in the expected punishment within a calibrated, life-cycle, general equilibrium model of crime choice. Their model is remarkably effective in accounting for the evolution of the US property crime rate over the last twenty-five years on the basis of changes in wage inequality, employment opportunities, age distribution and expected punishment. The focus of their analysis is positive. With the aim of accounting for changes in the crime rate, they take the education distribution parametrically and let it vary according to its evolution in the data over the relevant period. Our focus is more normative. For this reason, we endogenize investment in education and the marginal returns to education. We also distinguish between ability to earn and education, and look at how selection into education depends on ability. The aim is to compare the relative cost-effectiveness of early-intervention policies, such as education policies, with ex post forms of intervention, such as increased punishment. Cozzi (2004) also

uses a calibrated equilibrium model to investigate the extent to which differences in poverty and labor market opportunities can rationalize the higher crime rates among African-American males.

Imrohoroglu, Merlo, and Rupert (2000) endogenize police expenditure and the degree of redistribution by adding a political economy dimension to an equilibrium model of crime. Crime rates, police expenditure and redistribution are determined by the initial distribution of earning abilities and the apprehension technology as the joint outcome of majority voting. The model is able to account for several feature of US data, such as the positive correlation between police expenditure and redistribution and the lack of correlation between redistribution and crime.

Finally 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 and is thus necessarily static and partial equilibrium.

The structure of the paper is the following. Section 2 introduces the model. Section 3 discusses the estimation strategy while the calibration is discussed in Section 4. Section 5 simulates the model and discusses the effect of alternative policies. Section 6 concludes.

2 The Economic model

2.1 Environment

The model has an overlapping generation structure. Time is discrete.

Demographics: The economy is populated by a continuum of individuals. At each date a new cohort of unit mass starts life. We denote by $j \in J$ the age of an individual. Individuals are born at age zero, cannot work beyond the compulsory retirement age j_r and die at age $\bar{j}$. The conditional probability of surviving from age j to $j+1$ is λ_j and the unconditional probability of surviving up to age $j > 0$ is $\Lambda_j = \Pi_{s=0}^{j-1}\lambda_s$.[6]

Preferences: Preferences are given by

$$E\sum_{j=0}^{\bar{j}} \beta^j \Lambda_j u(c_j, i_j^s, d_j^c),$$

where c_j denotes consumption at age j. i_j^s and d_j^c are indicator functions. The former equals one if the individual is in education at age j while the latter assumes value one for an agent engaged in crime at

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

age j. They are zero otherwise. The felicity function is

$$u(c_j, i_j^s, d_j^c) = u(c_j) + \psi_j(\theta) i_j^s + \chi d_j^c = \frac{c_j^{1-\sigma}}{1-\sigma} + \psi_j(\theta) i_j^s + \chi d_j^c. \tag{1}$$

The parameters $\psi_j(\theta)$ captures the (dis)utility of the effort associated with education for a student with measured ability θ. The parameter χ captures the (dis)utility of engaging in crime other than the opportunity cost stemming from foregone market returns. We have chosen this specification since the flexibility it provides turns out to be crucial to match enrolment rates by ability and the (local) elasticity of crime to expected punishment observed in the data.

Agents do not value their offsprings' welfare and discount the future at rate β.

Education: Educational attainment e can take values in $E = \{e_0, e_1, e_2\}$.[7] To achieve education level e_n an agent has to be in school up to age j_n with $j_{n+1} > j_n$ and $j_0 = 0$. Hence ed, the number of years of education, takes values in $ED = \{0, 1, \ldots, j_2\}$. We denote by $i_j^s \in \{0, 1\}$ the choice to study or not at age $j \leq j_2$, with $i_j^s = 1$ if the individual is in education and zero otherwise. The direct per-period out of pocket cost (fee) of studying towards a degree e is f_e and the utility of being in education e_n, $\psi_j(\theta)$, is constant for any $j_{n-1} < j \leq j_n$ Students who start a course of study are assumed to be committed to it till its end. Agents who abandon education cannot go back to school at a later date. After completing school, agents enter the labour force. Market productivity increases only with the completion of an additional level of education.

Crime choice: Education and retirement are incompatible with crime but all agents in the labour force can choose to engage in criminal activities regardless of their employment status[8]. We denote by $d_j^c \in \{0, 1\}$ the choice to engage or not in crime at age $j < j_r$, with $d_j^c = 1$ if the individual engages in crime and zero otherwise. Criminal activity amounts to theft. In the current version of the model, only workers can be robbed, while students and pensioners cannot[9]. For a victim being the target of a theft involves losing a fraction α of post-tax labour income. This is equivalent to say that workers face a random, multiplicative shock v with support $\Psi = \{1 - \alpha, 1\}$ to their post-tax labour income. The probability of being the victim of a crime is $\pi_v = Pr(v = 1 - \alpha)$. For simplicity we assume criminals cannot target their victims and each of them obtains a fraction α of the average post-tax labour income.

[7] These correspond to "less than high school," "high school" and "at least college."

[8] This is the same assumption as in Imrohoroglu, Merlo, and Rupert (2004). It is consistent with evidence reported in Merlo (2001) that 71% of state prisoners in 1979 were employed prior to their conviction.

[9] This will be relaxed in a future draft.

An agent engaged in crime in the current period is apprehended and sent to jail with probability π_a. The length of the jail sentence is τ periods, starting from the current one. A convicted criminal keeps both her assets and the proceeds from her last crime, but she cannot access them while in jail. No optimising choice takes place while in jail and utility is exogenously given and equal to $\bar{u}$.

Endowments: Agents cannot hold a job while in education. Once they have left education and entered the labour market, the supply labour their unit labour endowment inelastically. An agent's labour supply is subject to a multiplicative i.i.d. employment shock equal to $0 < \underline{l} < 1$ with probability $\pi_u > 0$ and one with probability $1 - \pi_u$.[10]. Therefore the actual labour supply l can take values in $L = \{1, \underline{l}\}$.

The efficiency associated with an agent's labour supply l is

$$h_j(\theta, e) = \exp(\theta + \xi_j(e)), \tag{2}$$

where $\theta \in \Theta$ is an agent's innate level of ability. In each generation, the share of individuals with innate ability θ is $\eta(\theta)$ with $\int_\Theta \eta(\theta) = 1$.

Production technology: Firms are identical and use human and physical capital to produce output according to the production function

$$Q(H, K) = H^{1-\alpha} K^{\alpha}, \tag{3}$$

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

$$H = [\alpha H_1^{\rho} + \beta H_2^{\rho} + (1 - \alpha - \beta) H_3^{\rho}]^{\frac{1}{\rho}}. \tag{4}$$

of the stocks of human capital H_n associated with education level ed_n. Workers with the same level of educational attainement are perfect substitutes. Physical capital depreciates at the exogenous rate δ.

Market arrangements: Markets for factors of production and the final good are competitive. There are no state-contingent markets to insure against income risk, but workers can self-insure by saving into the risk-free asset. They also face an exogenous borrowing limit $a' \geq \bar{a}$, where a' denotes the stock of riskless asset at the beginning of the next period.. Assets of agents who die before age $\bar{j}$

[10]This way of modelling unemployment shocks is the same as in Heathcote, Storesletten, and Violante (2003). It allows us to contain the computational burden by using a time interval of a year in our calibration while still being able to accommodate a shorter unemployment duration. The latter is around one quarter for the US.

are distributed to the newborns in such a way that the cross-sectional distributions of wealth across deceased and newborn agents coincide[11].

We denote by $w(e)$ the wage per efficiency unit of labour of a worker with education e and by $\hat{r}$ the riskless interest rate.

Government: The government administers a pay-as-you-go pension system, the criminal justice system, spends on wasteful public expenditure and transfers and collects taxes. It balances the budget at all times.

In each period, it pays a pension benefit p to each pensioner and bears a total cost m for each convicted criminal. The government also pays a yearly subsidy $sub_e(a)$ to a student with wealth a studying for a degree leading to education level e. Both pension benefits and student subsidies are tax-exempt while labour and capital income are taxed at proportional rates t_l and t_k respectively.

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.

Timing: The timing of events is as follows. At the beginning of each period potential students decide whether to enter the labour market in the current period and all workers draw their labour supply and decide whether to engage in criminal activity or not. At this point criminals may be arrested. At the end of period agents receive their labour income or transfers and decide how much to consume and save[12].

2.2 Recursive Representation

The consumers' optimization problem admits a recursive representation (see Gallipoli (2005) for proofs). The individual state is fully characterized by age j, the worker's type associated with innate ability θ, completed years of education e, beginning-of-period asset holdings a, the labour supply realization l and the victimization shock v.

Let E_z denote the expectation operator with respect to the probability distribution of z. Let us also denote by $r = (1 - t_k)\hat{r}$ the post-tax interest rate equals and by $y_j(\theta, e, l)$ an agent's flow of consumable resource other than financial income. Let the superscripts s, nc, na, a, pr and r index respectively students, non-criminal, non-apprehended criminal, criminals apprehended in the current period, agents already in prison at the beginning of the period and pensioners. Then disposable non-financial "income"

[11] The details of the mechanism generating bequests are discussed in Section 4.

[12] Apprehended criminals do not supply any labour in the market. Payment of labour income, net of any losses due to crime, are paid at the end of each period before optimal consumption levels are chosen.

for agent i is given by

$$y_j^i(\theta, e, l, v) = \begin{cases} sub_e(a) - D_e & \text{if } i = s \\ (1-t_l)(1-v)\, w(e)\, h_j(\theta, e) & \text{if } i = nc \\ (1-t_l)(1-v)\, w(e)\, h_j(\theta, e) + \alpha\bar{y} & \text{if } i = na \\ \alpha\bar{y} & \text{if } i = a \\ 0 & \text{if } i = pr \\ p & \text{if } i = r. \end{cases} \tag{5}$$

Students and pensioners do not pay taxes on their flow of consumable resources. Agents engaged in crime in the current period receive their current illegal income $\alpha\bar{y}$. If apprehended they go to jail before receiving their labour income labour. Agents who are not in jail in the current period receive their labour income net of taxes but can be robbed of a share v of it. The associated dynamic budget identity for individual i is

$$a' = \begin{cases} a(1+r) + y^i - c & \text{if } i \neq pr \\ a(1+r) + y^i & \text{if } i = pr. \end{cases}$$

Given the borrowing constraint and the lack of a bequest motive, the individual maximization problem is also subject to the constraints

$$a' \geq \bar{a},\ a'_{\bar{j}} \geq 0,\ a_0 \text{ given} \tag{6}$$

The value function of a current student satisfies the Bellman equation

$$V_j^s(\theta, e_n, a) = \max_c u(c) + \psi_j(\theta) + \beta\lambda_j V_{j+1}^s(\theta, e_n, a') \tag{7}$$

if $j_n < j < j_{n+1}$ and

$$V_j^s(\theta, e_n, a_j) = \max_c u(c) + \psi_j(\theta) + \beta\lambda_j \max\{V_{j+1}^s(\theta, e_{n+1}, a'), E_l V_{j+1}^w(\theta, e_{n+1}, a', l)\} \tag{8}$$

otherwise. The superscripts s and w index respectively students and workers. A student chooses consumption optimally subject to her budget constraint. Furthermore, in the year following the completion of its current education course the student has to choose optimally whether to study further or enter the

labour force before knowing her current labour supply realization. Since the highest attainable degree is college, the student's problem is subject to the terminal condition $V_{j_2}^{s}(\theta, e, a) = E_l V_{j_2}^{w}(\theta, e, a, l)$.

If one denotes by V^{nc} the value function of a worker not engaged in crime and by V^{c} the value function of a criminal *gross* of the cost of crime χ, the problem of a labour force participant can be written as

$$V_j^{w}(\theta, e, a, l) = \max\left\{V_j^{c}(\theta, e, a, l) + \chi, V_j^{nc}(\theta, e, a, l)\right\}. \tag{9}$$

After observing her labour supply realization the agent chooses whether to engage or not in crime. In the former case she is apprehended with probability α and her value function is

$$V_j^{c}(\theta, e, a, l) = (1 - \pi_a) E_v V_j^{na}(\theta, e, a, l) + \pi_a V_j^{a}(\theta, e, a, l).$$

The lifetime expected utility of a criminal entering jail[13] is

$$V_j^{a}(\theta, e, a, l) = \frac{1}{\Lambda_j}\left[\bar{u}\sum\nolimits_{s=0}^{\tau-1}\beta^s\Lambda_{s+j} + \beta^{\tau}\Lambda_{j+\tau}E_l V_{j+\tau}^{w}(\theta, e, a_{j+\tau}, l)\right]. \tag{10}$$

Finally, the value function of an agent who is out of jail in the current period - i.e. $i = nc, na$ - is given by

$$V_j^{i}(\theta, e, a, l) = E_v\left[\max_c u(c) + \beta\lambda_j E_l V_{j+1}^{w}(\theta, e, a', l')\right]. \tag{11}$$

The agent is subject to the random shock v associated with being robbed and is uncertain abour her next-period labour supply realization.

2.3 Stationary Equilibrium

The equilibrium concept we use is that of recursive, stationary, competitive equilibrium following Stokey, Lucas, and Prescott (1989). To streamline notation we denote by $s \in S$ the vector of state variables $(\theta, e, a, l) \in \Theta \times E \times A \times L$ and denote by capital letters aggregates of individual quantities denoted by the corresponding small case letter. With some abuse of notation, we use integrals even when summing

[13] Note that a model period is one year. When the prison term is longer than a year we assume that an apprehended criminal receives utility $\bar{u}$ for the entire year in which she enters jail and the remaining fraction of the following year. After leaving jail her labour supply is reduced according to the remaining fraction of the year and is subject to the same multiplicative shock as any other worker.

over discrete variables.

Definition 1 *For a given set of government policies* $\{\tau, p, G, sub_e(a), t_l, t_k\}$ *and an apprehension probability* π_a*, a recursive stationary equilibrium is a collection of value functions* V_j^i*, individual decision rules* $\{i_j^s, d_j^c\} : S \to \{0,1\}$ *and* $\{c_j^i, a_j'^i\} : S \times \Psi \to R$*, decision rules* $\{K, H_n\}$ *for firms, prices* $\{r, w_{e_n}\}$*, time-invariant measures* $\mu_j^i : S \to [0,1]$*, a victimization probability* π_v *and an average labour income* $\bar{y}$ *such that:*

1. *Given* $\{r, w_n\}$, $\{c_j^i(s,v), a_j'^i(s,v), i_j^s(s), d_j^c(s)\}$ *for* $i \neq pr$ *solve the set of problems (7)-(11) and* V_j^i *are the associated value functions. Moreover,* $a_j'^{pr}(s,v) = (1+r)\,a$.

2. *Given* $\{r, w_{e_n}\}$ *,* K *and* H_n *satisfy*

$$r + \delta = F_K \tag{12}$$

and

$$w(e_n) = F_{H_n}. \tag{13}$$

3. *Factor and product markets clear*[14] *or*

$$H_n = \int_{J \times [S|E]} lh_j(\theta, e_n)\, d\mu_j^{na}(\theta, e_n, a, l) \tag{14}$$

and

$$Q - \delta K = C + M + G + F.$$

4. *The government budget is balanced*

$$M + G + P + SUB = t_k \hat{r} K + t_l \int_{J \times S} w(e_s) lh_j(\theta, e_s)\, d\mu_j^{na}(s) \tag{15}$$

5. *The victimization rate coincides with the crime rate and satisfies*

$$\pi_v = \left(\int_{J \times S} d\mu_j^w(s) \right)^{-1} \int_{J \times S} d\mu_j^c(s). \tag{16}$$

[14] By Walras law, market clearing on the good market and the markets for the three types of labour ensures that the capital market clears.

6. The average disposable labor income of employed workers satisfies

$$\bar{y} = \left(\int_{J \times S} d\mu_j^{na}(s) \right)^{-1} \int_{J \times S} w(e) l h_j \left(\theta, e_s \right) d\mu_j^{na}(s). \tag{17}$$

3 Estimation

The parameters of the model are obtained by a combination of estimation and calibration using data from the US. We estimate components of the wage process and the aggregate production function.

3.1 Estimating wage equations: skill prices variation and age profiles

An important characteristic of the model is that the three types of human capital represent different inputs to the production function, not necessarily perfectly substitutable and may have relative prices that vary over time in response to changes in either supply or demand for skills.

So as to be able to simulate our model, we need to have a distribution of unobserved heterogeneity affecting wages and education choices. In the 1972 wave of the PSID several IQ measures were elicited for households heads and after some examination one of them was deemed to be the most accurate and released. We use the cross-sectional distribution of such IQ test scores to approximate the permanent heterogeneity in our sample.

In figure (3.1) we report the measured IQ densities for the whole 1972 sample and a selected sub-sample based on our criterion. It seems that IQ density exhibits a long left tail.

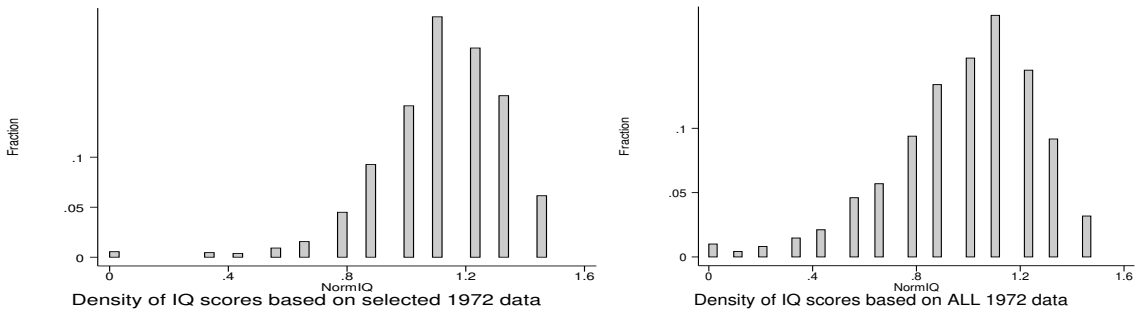

Figure 1: Density of IQ measurement from 1972 PSID wave, for the whole sample and a comparable sub-sample.

Permanent characteristics are only one of the determinants of wages and other dimensions of heterogeneity must be analysed. We start by specifying an education specific wage equation for individual

i's (log) wages in period t

$$\ln w_{eit} = w_{et} + g_e\left(age_{it}\right) + u_{eit} \tag{18}$$

where w_{et} represents the log of the aggregate price of human capital for education group e and where $g_e\left(age_{it}\right)$ is the education specific profile of wages. The unobservable component u_{eit} is specified to be

$$u_{eit} = b_e(\theta_i) + m_{it} \tag{19}$$

where $b_e(\theta_i)$ is a function of unobserved fixed effects (ability) and m_{it} is measurement error, assumed *iid*. Self-selection implies that fixed effects are correlated with both education decisions and observed wage rates. We use cross-sectional variation to identify the gradients of age and ability by estimating the following equation on data from the 1972 wave of the PSID separately for each education group

$$\ln w_{ei} = c^{edu} + \beta^{edu} IQ + \alpha_1 age + \alpha_2 age^2 + \varepsilon_{ei} \tag{20}$$

where IQ denotes an individual test score, c^{edu} is a constant and $\varepsilon_{ei} = m_{it}$.

If we assume that $g_e\left(age_{it}\right)$ be a polynomial in age such that $g_e\left(age_{it}\right) = \alpha_0^e + \alpha_1^e age_{it} + \alpha_2^e age_{it}^2$, it follows that the intercepts of the 3 education specific equations, c^{edu}, estimate the sum of the age profile component α_0^e and the education specific price w_{e1972}. Some normalization assumption is necessary to disentangle these 2 components. The method we use in order to normalize the α_0^e terms (and therefore the skill prices) is described in more detail when we discuss estimation of aggregate technology parameters. The quadratic age profiles are used in the numerical simulations[15].

In order to identify time variation in skill prices we exploit the panel dimension of the PSID data set. Using equation (18), we can identify the year-specific (log) changes in wage growth for each education group by looking at individual (log) wage changes. We acknowledge that the age composition in each of our education subsamples is different. We control for different average age in each education group by estimating a first step regression of log real hourly earnings on age and a constant

$$\ln w_{eit} = \kappa_e + \lambda age_{eit} + residual_{eit}$$

[15]In estimating age profiles from the 1972 cross-sectional data we ignore cohort effects, which are likely to induce a downward sloping profile at older ages. At the moment we do not address this issue.

The residual can be interpreted as the log real hourly earnings of an agent after controlling for the group age. We then define the first difference of the residuals as

$$\eta_{eit} = residual_{eit} - residual_{eit-1}$$

and identify the growth rates of wages in different groups by estimating

$$\eta_{ei,year} = dummy_{\{year\}} + \epsilon_{ei,year} \tag{21}$$

for all years between 1968 and 1997. Standard errors are robust and use cluster adjustment[16].

3.2 Wage data and results

For the estimation of wage equations we use cross-sectional data from the 1972 PSID wave. We do not use individuals associated with the Census low income sample, the Latino sample or the New Immigrant sample and focus instead on the SRC core sample, which did not suffer any systematic additions or reductions between 1968 and 2001 and was originally representative of the US population. We drop people with a zero test score because most of them did not take the test seriously enough to be part of the sample.

The main earnings' variable in the PSID refers to the head of the household[17] and is described as total labor income of the head[18]. We use this measure, deflated into 1992 dollars by the CPI-U for all urban consumers. By selecting only heads of household we ignore other potential earners in a family unit and restrict our attention to people with relatively strong attachment to the labor force. We include both men and women as well as whites and non-whites[19].

Information on the highest grade completed is used to allocate individuals to three education groups: high school drop-outs (LTHS), high school graduates (HSG) and college graduates (CG). In fact, the maximum age in the cross-section turns out to be 62. The non constant terms of the age polynomials from the wage equation are presented in table (1).

[16] No constant is estimated in this equation.

[17] In the PSID the head of the household is a male whenever there is a cohabiting male/female couple. Women are considered heads of household only when living on their own. We do not address the related sample issues explicitly, but any gender effects are likely to be captured in the ability estimates.

[18] This includes the labor part of both farm and business income, wages, bonuses, overtime, commissions, professional practice and others. Labor earnings data are retrospective, as the questions refer to previous year's earnings, which means that 1968 data refer to 1967 earnings.

[19] There is evidence (see Cozzi (2004)) that property crime is done mostly by males. We do not exclude females from our sample in order to keep consistency between our CPS and PSID data sets.

Table 1: Results of the estimation for the education specific cross-sectional equations.

Dependent variable: log hourly earnings		
coeff.	point estimate	S.E.
	Education=LTHS	
constant	.2185943	.3889016
IQ	.3271764	.1317296
age	.0951891	.0189531
age^2	-.0011132	.0002244
	Education=HSG	
constant	.5775383	.2269027
IQ	.3332425	.0908616
age	.0787976	.0110038
age^2	-.0008399	.0001401
	Education=CG	
constant	-.1005019	.3833756
IQ	.0387147	.1597967
age	.1396539	.0176611
age^2	-.0014817	.0002271

Figure (2) plots the age profiles (in logs) implied by the polynomial estimates for different education groups under the assumption that the constant terms are zero.

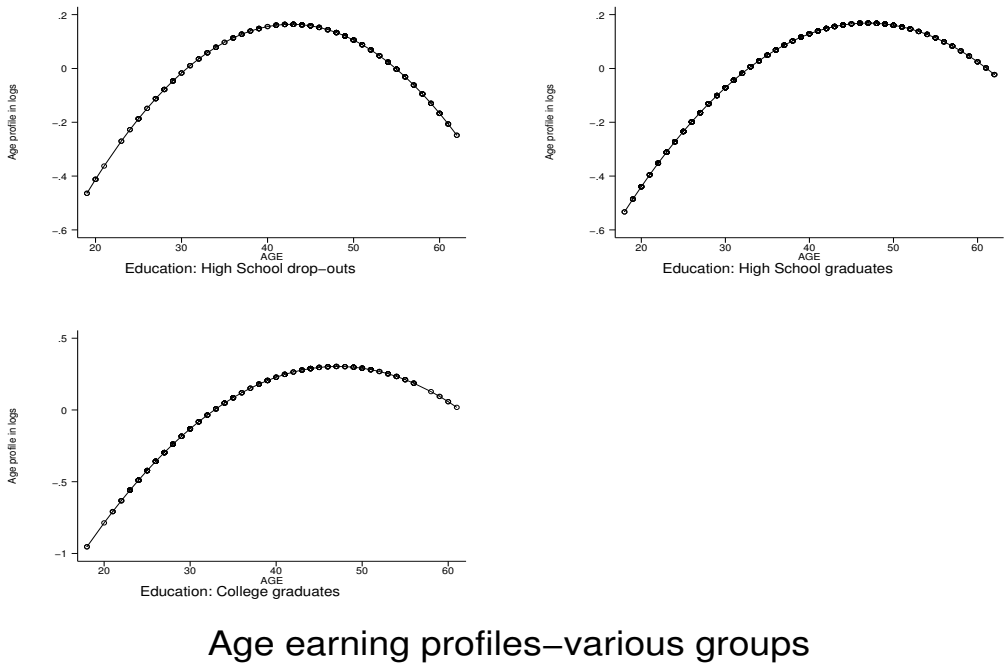

Figure 2: (Log) Age profiles of labor efficiency by education group - age on the horizontal axis.

A detailed description of our sample selection for the estimation of log changes in HC prices is reported in the appendix: in brief, we select heads of household aged 25-65 who are not self-employed and have positive labor income for at least 8 (possibly non continuous) years. The estimated log changes of price effects w_{et} for different education groups are presented in figure (3).

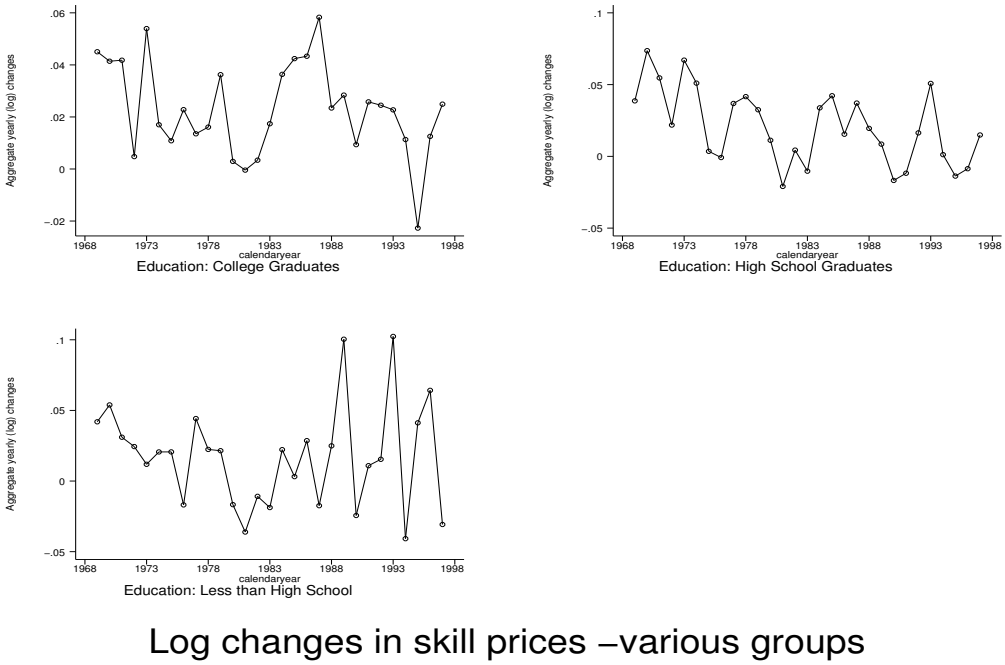

Figure 3: Estimated log changes of marginal labor productivity, by education and year.

The growth rates of skills prices are of fundamental importance to help identify human capital aggregates and the parameters of aggregate technology.

3.3 Using CPS to obtain data for the aggregate production function.

Estimation of the aggregate production function requires the total wage bills for each of the education groups, and in the general CES case we also need measures of human capital in each of these groups. We use the March supplement of the Current Population Survey (CPS) to obtain these. The CPS is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics.[20] The wage bills are straightforward to obtain. We just add up the earnings of each of the three education groups and then scale up the figures to match the entire US economy.

When we need to estimate a CES production function the issue is more involved because we also need to estimate the *quantity of human capital* in each year. To achieve this we need an aggregate price series for each of the education groups; our estimates from the PSID provide the growth of prices over time, but a normalisation assumption on each price is necessary. Any normalisation will correspond to a set of relative prices at a given point in time. However, we still have one degree of freedom: in fact, after setting the initial relative price of high school and of college graduate labour we can choose the

[20]The survey has been conducted for more than 50 years. Statistics on the employment status of the population and related data are compiled by the Bureau Labor Statistics (BLS) using data from the Current Population Survey (CPS).

utility costs of education to match the proportions going into each of the educational categories. In other words with unobserved costs the data can be rationalised either with high returns and high costs or low returns and low costs. The particular normalisation we choose will not affect the simulation of the policy changes.

The adult universe (i.e., population of marriageable age) is comprised of persons 15 years old and over for March supplement data and for CPS labor force data. Each household and person has a weight that we use in producing population-level statistics. The weight reflects the probability sampling process and estimation procedures designed to account for nonresponse and undercoverage.

We use the CPI for all urban consumer (with base year 1992) to deflate the CPS earnings data and drop all observations that have missing or zero earnings. Since the earning data are top-coded for confidentiality issues, we have extrapolated the average of the top-coded values by using a tail approximations based on a Pareto distribution.[21]

Figure (4) reports the number of people working in each year by education group, as reported by the CPS.

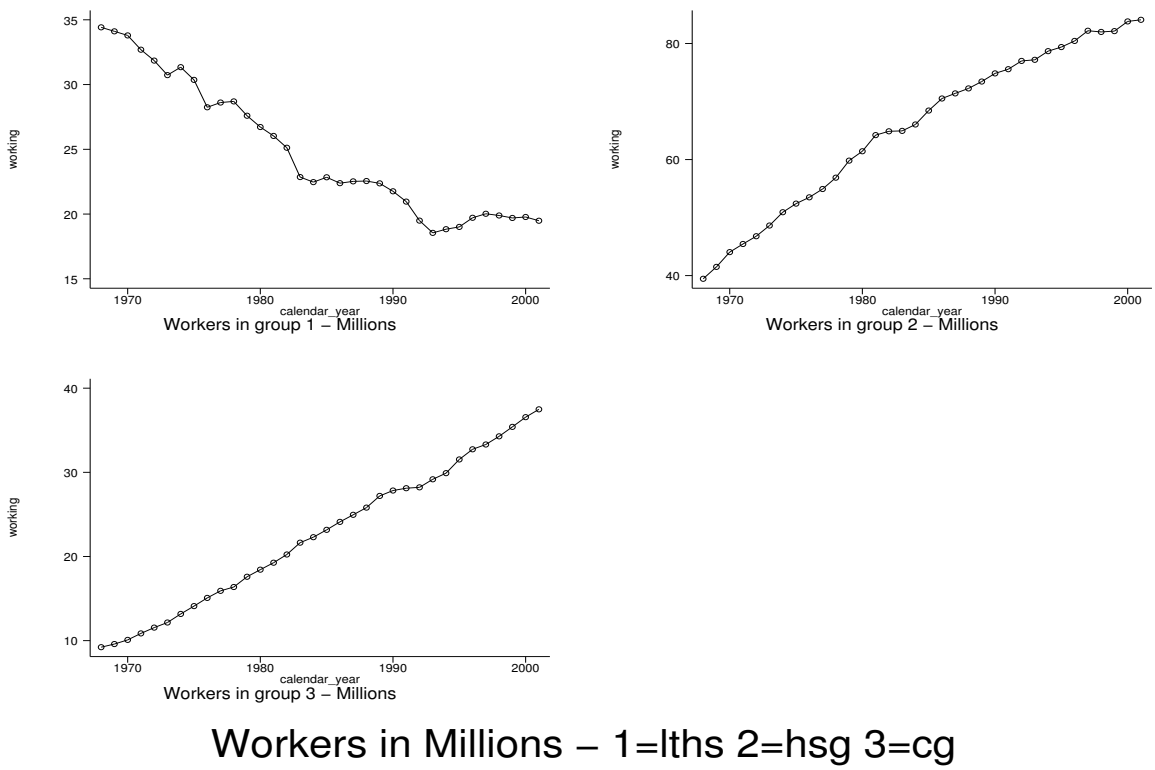

Figure 4: Employed workers in millions, by education and year

[21] This procedure is based on a general approach to inference about the tail of a distribution originally developed by Hill (1975). This approach has been proposed as an effective way to approximate the mean of top-coded CPS earning data by West (1985); Polivka (2000) provides evidence that this method closely approximates the average of the top-coded tails by validating the fitted data through undisclosed and confidential non top-coded data available only at the BLS.

It is clear that some strong and persistent trends towards higher levels of education have characterized the sample period.

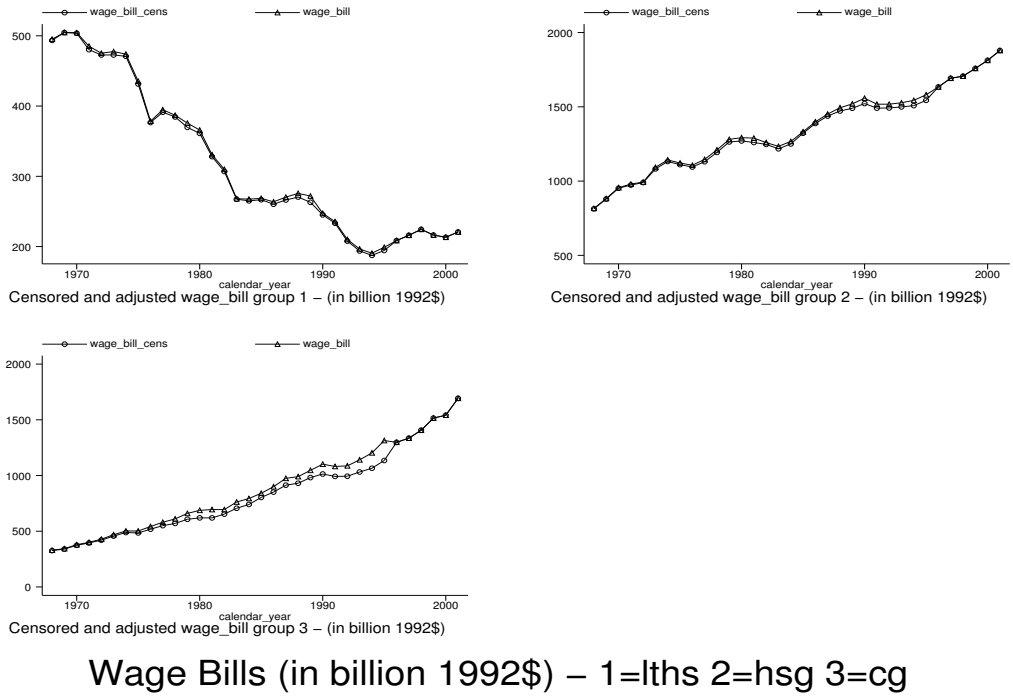

Figure 5: Total earned labor income, by education and year. Total in billions of 1992 dollars.

Figures (3.3 – 3.3) plot both the average earnings by year and total wage bills in billions of dollars for the 3 education groups (1=less than High School, 2=High School finishers, 3=College graduates). Since CPS earning data until 1996 are top coded we report both the censored mean and a mean adjusted by using a method suggested by the BLS (see West (1985)) which is based on the original Hill's estimator to approximate exponential tails. The difference between the two averages is larger for the most educated people who tend to be more affected by top-coding. We include also self-employed people in the computation of these aggregates; however, their exclusion has almost no effect on the value of the wage bills and human capital aggregate, as they never represent more than 5% of the working population in a given education group (and most of the time much less than that).

Given a series of log prices for the HC aggregates, it is possible to divide the wage bills by the exponentiated value of such prices (whose dynamic behaviour is estimated using PSID data) and finally obtain point estimates of the value of efficiency weighted total labor supply (human capital aggregates) by education and year. Our estimates of human capital aggregates are discussed in the next section, since the specific normalization we use has an effect on the estimation of some technology parameters.

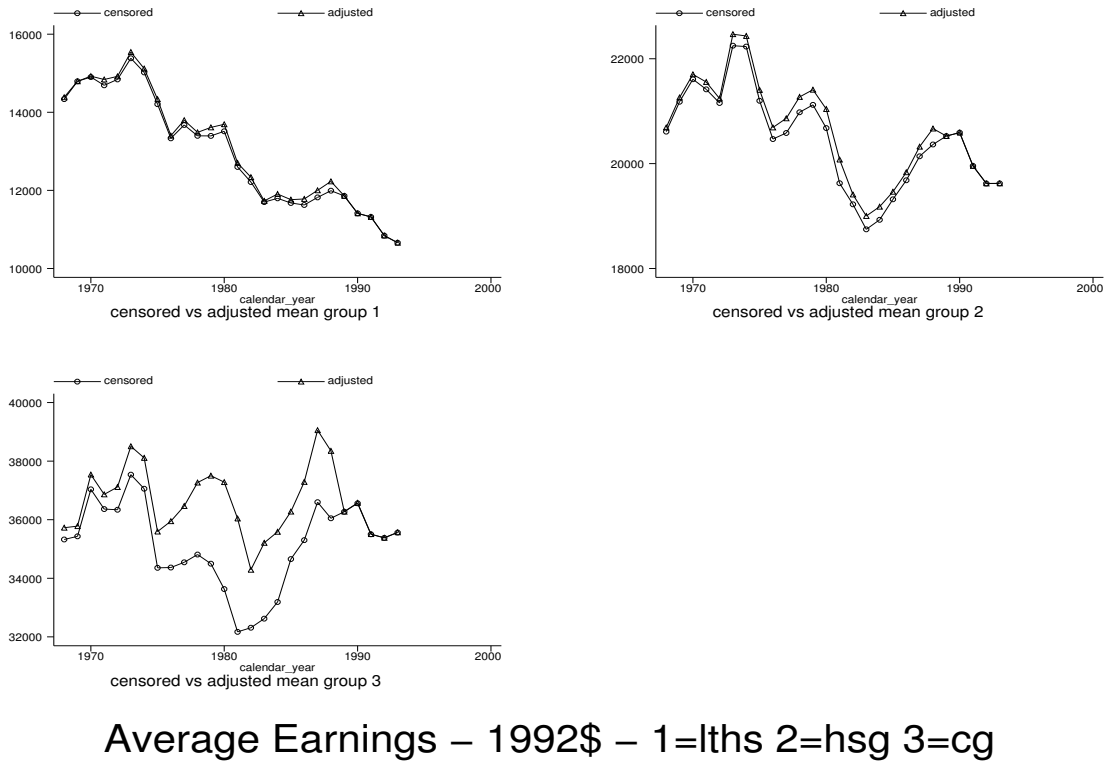

Figure 6: Average earned labor income, by education and year. Average in units of 1992 dollars.

3.4 Aggregate Production Function

In estimating technology parameters, we start from the relatively easier case of Cobb-Douglas technology. Let aggregate output Y be produced through the following technology

$$Y = \left(H_1^a H_2^{(1-a)b} H_3^{(1-a)(1-b)} \right)^{1-\alpha} K^{\alpha}$$

Using NIPA data we find the share of capital α to be between 0.3 and 0.35, depending on whether we correct for housing stocks. The share parameters a and b can be easily expressed as a function of the aggregate wage bills. If we apply this procedure separately for each year we can pinpoint the evolution of these functions over the sample period.

Figure (3.4) reports the value of the share parameters (with bounds equal to 2 standard errors) for the shares associated to each human capital variety. In figure (3.4) the line that is increasing over the sample periods represents the College graduates share $(1-a)(1-b)$, whereas the downward sloping one represents the high school drop-outs share a. The almost flat line on top is the high school graduates share $(1-a)\,b$.

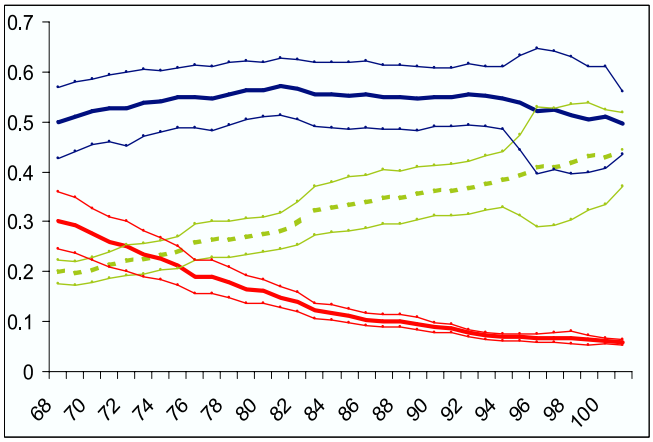

Figure 7: Labor shares in human capital input of technology, computed using Cobb-Douglas specification (with bounds equal to +/- 2 standard errors). Period: 1968-2000. Larger bounds after 1996 are due to changes in top-coding of income in the CPS.

The time averages of such shares are 0.33 for college graduates, 0.54 for high school graduates and 0.14 for high school dropouts. The evolution of the college graduates labor share over time more than doubles (from 0.2 to 0.4) whereas the share of less-than-high-school labor falls dramatically from over 0.3 to roughly 0.06. These findings, together with our PSID estimates of changes in the marginal products of labor, seem to confirm that major shifts in technology have taken place between the late 1960s and the end of the century.

We now proceed to relax the restriction of unit elasticity implicit in the Cobb-Douglas specification but keep the isoelasticity assumption. We write the aggregate technology as

$$Y = K^{\alpha} H^{1-\alpha}$$

$$H = \left(aH_1^{\rho} + bH_2^{\rho} + (1-a-b)\, H_3^{\rho}\right)^{\frac{1}{\rho}}$$

The ratio of marginal products of different human capital inputs ($MPHC^{edu}$) can now be written as

$$MPHC^3/MPHC^1 = \frac{(1-a-b)}{a}\frac{H_3^{\rho-1}}{H_1^{\rho-1}}$$

$$MPHC^3/MPHC^2 = \frac{(1-a-b)}{b}\frac{H_3^{\rho-1}}{H_2^{\rho-1}}$$

$$MPHC^2/MPHC^1 = \frac{b}{a}\frac{H_2^{\rho-1}}{H_1^{\rho-1}}$$

and terms can be rearranged to obtain expressions involving wage bills, like

$$WB^3/WB^1 = \frac{(1-a-b)}{a}\left(\frac{H_3}{H_1}\right)^{\rho} \tag{22}$$

$$WB^3/WB^2 = \frac{(1-a-b)}{b}\left(\frac{H_3}{H_2}\right)^{\rho} \tag{23}$$

$$WB^2/WB^1 = \frac{b}{a}\left(\frac{H_2}{H_1}\right)^{\rho} \tag{24}$$

which can be log-linearised and used to obtain estimates of ρ and ratios of share parameters. At this point the normalisation of skills prices becomes important, as it determines the relative sizes of the human capital aggregates H_{edu}.

Remember from equation (20) that the marginal products of human capital types in the cross-sectional equation based on the 1972 PSID wave cannot be identified separately from the intercept of the (log) age profiles α_0^{edu} of the age polynomial. In other words, the amount of log hourly wage that is attributed to $\log MPHC^{edu}$ cannot be distinguished from the amount attributed to a component of the age polynomial. Therefore a normalizing assumption is needed to disentangle these two components of wages.

Any normalizing restrictions on the log of the marginal products of each education type ($MPHC^{edu}$) have an effect on the estimates of the share parameters of different aggregate human capital types in technology. This can be easily seen by way of example. We know that

$$\log\left(WB^2/WB^1\right) = \log\left(\frac{b}{a}\right) + \rho\log\left(\frac{H_2}{H_1}\right)$$

where $\frac{H_2}{H_1} = \frac{\exp\left(\alpha_0^2\right)}{\exp\left(\alpha_0^1\right)}\frac{\tilde{H}_2}{\tilde{H}_1}$ where $\frac{\tilde{H}_2}{\tilde{H}_1}$ are the ratio of human capital aggregates obtained under the assumption that both α_0^1 and α_0^2 are equal to zero (that is when the constant term κ^{edu} in the cross-sectional wage equation is fully attributed to the marginal product of human capital). Then we can write

$$\log\left(WB^2/WB^1\right) = \left[\log\left(\frac{b}{a}\right) + \rho\log\frac{\exp\left(\alpha_0^2\right)}{\exp\left(\alpha_0^1\right)}\right] + \rho\log\left(\frac{\tilde{H}_2}{\tilde{H}_1}\right) \tag{25}$$

Notice that the ρ parameter in equation (25) is identified under any rescalings of the ratio $\frac{H_2}{H_1}$, because the log transformation captures any rescaling factor in the constant term. This means that the estimation of ρ does not change with alternative normalisations of the α_0^{edu} terms. In a similar way, we can

decompose the ratios WB^3/WB^1 and WB^3/WB^2 and then define a system of linear equations like

$$\log\left(\frac{b}{a}\right) + \rho \log \frac{\exp\left(\alpha_0^2\right)}{\exp\left(\alpha_0^1\right)} = X$$
$$\log\left(\frac{1-a-b}{a}\right) + \rho \log \frac{\exp\left(\alpha_0^3\right)}{\exp\left(\alpha_0^1\right)} = Y$$
$$\log\left(\frac{1-a-b}{b}\right) + \rho \log \frac{\exp\left(\alpha_0^3\right)}{\exp\left(\alpha_0^2\right)} = Z$$

where (X, Y, Z) is a vector containing the estimates of the constant terms from the log-linearised estimation of equations $(22-24)$.

The above system can provide an estimate of the shares a,b and (1-a-b) as a function of the normalization chosen for the α_0^{edu} terms. This is linked to the human capital aggregates, which also change with the α_0^{edu} terms. An easier notation to write the linear system above is

$$\delta_2 - \delta_1 + \rho\left(\alpha_0^2 - \alpha_0^1\right) = X \tag{26}$$

$$\delta_3 - \delta_1 + \rho\left(\alpha_0^3 - \alpha_0^1\right) = Y \tag{27}$$

$$\delta_3 - \delta_2 + \rho\left(\alpha_0^3 - \alpha_0^2\right) = Z \tag{28}$$

with $\delta_1 = \log(a)$, $\delta_2 = \log(b)$ and $\delta_3 = \log(1-a-b)$. The system in matrix form is

$$\begin{bmatrix} 1 & -1 & 0 & -\rho & \rho & 0 \\ 0 & -1 & 1 & -\rho & 0 & \rho \\ -1 & 0 & 1 & 0 & -\rho & \rho \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \alpha_0^1 \\ \alpha_0^2 \\ \alpha_0^3 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

or

$$Ax = B$$

and at least a solution to this system exists if and only if $rank\,[A|B] = rank\,[A]$. Therefore we have to check whether the vector (X, Y, Z) is such that, if we set $\left(\alpha_0^1, \alpha_0^2, \alpha_0^3\right)$ to some arbitrary (normalising) values, we can solve the system above for a triplet $(\delta_1, \delta_2, \delta_3)$ without having any contraddicting solutions:

Dependent Variable: log of wage bill ratio			
coeff.	point estimate	S.E.	
	Wage bills: Edu 3 / Edu 2		obs.=26
ρ	.3952441	.0912868	
trend	.0192798	.0018233	
constant	-.8612072	.0282592	
	Wage bills: Edu 3 / Edu 1		obs.=26
ρ	.5967261	.2664686	
trend	.0393077	.0193558	
constant	-.1914386	.1129425	
	Wage bills: Edu 2 / Edu 1		obs.=26
ρ	.3354899	.2949902	
trend	.0377996	.0157973	
constant	.5802938	.0456587	

Table 2: Estimates of unrestricted, log-linearised wage bills ratios equations.

given that the matrix A is of rank 2 (easily checked!) this will be true if and only if $X = Y - Z$ (to check this, just subtract the third equation from the second), which means that we have a system in 2 equations and 2 unknowns. The restriction on the technology shares $\exp(\delta_1) + \exp(\delta_2) + \exp(\delta_3) = 1$ guarantees that we can find all share parameters. Notice that not all the (X, Y, Z) triplets can guarantee existence of a solution for this system.

3.5 Estimation results for production function parameters

We estimate a version of equation (25) augmented by a linear time trend for each of the 3 wage bill ratios. The time varying regressor $\log\left(\frac{\tilde{H}_{2t}}{\tilde{H}_{1t}}\right)$ is based on the assumption that the terms c^{edu} in equations (20) identify the relative prices in 1972. To control for possible endogeneity of the human capital inputs in the production function, we adopt an IV approach with lagged regressors (lags from 1 to 4 periods back are included in the first step). The results of this specifications, separately estimated, are reported in table (3.5) with standard errors in parenthesis.

Using a joint estimation approach we are also able to test whether the estimates of the ρ parameters provided by different ratios are statistically different from each other (i.e. whether we can reject the isoelastic assumption)[22]. The results of such tests are reported in table (3.5).

The tests for equality of the ρ parameters are unable to reject the null hypothesis that the aggregate technology is isoelastic. We therefore estimate a restricted version of equations (20) in which we restrict the ρ to be the same for all ratios. The results for this speecification are reported in table (3.5).

Our restricted estimate for ρ is approximately .45 which corresponds to an elasticity of substitution

[22]Using a relatively short time series data set implies relatively large standard errors in our unrestricted estimation, with the noticeable exception of the edu3/edu2 ratio.

Testing the isoelastic restrictions

Estimates of ρ being tested (by wage ratios)	F-statistic	Prob. > F-statistic
$\rho_{(2/1)} = \rho_{(3/2)}$	0.03	0.8596
$\rho_{(3/2)} = \rho_{(3/1)}$	0.42	0.5208
$\rho_{(2/1)} = \rho_{(3/1)}$	0.61	0.4383
$\rho_{(2/1)} = \rho_{(3/1)} = \rho_{(3/1)}$	0.36	0.7019

Table 3: Tests for equality of elasticities of substitution among human capital inputs

Dependent Variable: log of wage bill ratio			
coeff.	point estimate	S.E.	
	Wage bills: Edu 3 / Edu 2		obs.=78
ρ	.452508	.1354589	
trend	.0181865	.002907	
constant	-.8446752	.0465645	
	Wage bills: Edu 3 / Edu 1		obs.=78
ρ	.452508	.1354589	
trend	.0497485	.0098961	
constant	-.2503577	.0608398	
	Wage bills: Edu 2 / Edu 1		obs.=78
ρ	.452508	.1354589	
trend	.031562	.0073416	
constant	.5943174	.0300404	

Table 4: Estimates of restricted, log-linearised wage bills ratios equations.

of around 1.8. Using a simple skilled/unskilled classification Katz and Murphy estimate the elasticity of substitution in production to be 1.41 with a standard error of .150. Heckman, Lochner and Taber (1998a) report a favorite estimate of the elasticity of substitution between skilled and unskilled equal to 1.441, whereas Johnson (1970) has an old estimate equal to 1.50. Notice that our elasticity estimate provides a measure of substitutability between 3 different types of workers, rather than two simple skill groups.

We also estimate a specification with only lags larger than 1 to instrument for endogeneity of human capital aggregates[23]: this gives a restricted ρ =.354822 (S.E..1974259), which implies a lower elasticity of 1.55, much closer to the skilled/unskilled estimates of the previous literature.

The share parameters for the CES production function can be identified by using the constants estimated from the wage bill ratio equations above. However, one must be careful what normalizing assumptions are made on the values of the α_0^{edu}terms in equation (25).

After experimenting with different alternatives, we have decided to set the intercept $\alpha_0^{edu}$of the log age profiles to values such that the *levels* of the age profiles for different education groups have all the same average over age. If we let $M^{edu} = \sum_{J=1}^{W_{\max}} \exp\left(\alpha_1^{edu} j + \alpha_2^{edu} j^2\right) / W_{\max}$, where $W_{\max}$is the

[23]This would control for potential error correlation up to lag 1.

Values of normalising $\left(\alpha_0^1, \alpha_0^2, \alpha_0^3\right)$ used in benchmark		
$\alpha_0^{edu=1}$	$\alpha_0^{edu=2}$	$\alpha_0^{edu=3}$
0	0.159	-1.164

Table 5: Value of the intercept in the (log) age profile polynomial. These values pinpoint human capital prices and technology shares.

maximum working age, the normalising assumption we are imposing is

$$\exp\left(\alpha_0^1\right) M^1 = \exp\left(\alpha_0^2\right) M^2 = \exp\left(\alpha_0^3\right) M^3 \tag{29}$$

This means that we find the α_0^{edu} such that

$$\alpha_0^1 - \alpha_0^2 = \ln\left(M^2\right) - \ln\left(M^1\right) \tag{30}$$

$$\alpha_0^2 - \alpha_0^2 = \ln\left(M^3\right) - \ln\left(M^2\right) \tag{31}$$

which can be solved by setting the value of one of the α_0^{edu} terms to some arbitrary value and solving equations $(30-31)$ for the remaining α_0^{edu} terms[24]. The resulting values are reported in table (3.5).

We plug these values from table (3.5) in the system $(26-28)$ together with the constants estimated in the wage bills equations (X=.5943174,Y=-.2503577,Z=-.8446752) reported in table (3.5). Notice that $X \approx Y - Z$ (easily check that $X - Y + Z = 5943174 + .2503577 - .8446752 \approx 0$) and we know that a solution exists. The system in levels (not logs) can be written as}

$$\begin{aligned}
\frac{b}{a} &= A = \exp\left[X - \rho\left(\alpha_0^2 - \alpha_0^1\right)\right] \\
\frac{1-a-b}{a} &= B = \exp\left[Y - \rho\left(\alpha_0^3 - \alpha_0^1\right)\right] \\
\frac{1-a-b}{b} &= C = \exp\left[Z - \rho\left(\alpha_0^3 - \alpha_0^2\right)\right]
\end{aligned}$$

and solving for a and b we get that $a = \frac{1}{1+B+A}$ =0.24, $b = \frac{A}{1+B+A}$ =0.41 and $(1-a-b)$ =0.35.

Notice that after choosing the α_0^{edu} terms in the age profile, we obtain the relative prices of different human capital in 1972, by using the constant terms c^{edu} estimated in (20). This allows to identify also a

[24]The solution of this system is:

$$\begin{aligned}
\alpha_0^1 &= 0 \\
\alpha_0^2 &= -\ln\ M^2\ + \ln\ M^1\ + \alpha_0^1 \\
\alpha_0^3 &= -\ln\ M^3\ + \ln\ M^1\ + \alpha_0^1
\end{aligned}$$

Normalised prices of human capital by education group in 1972				
	α_0	*constant*	*1972 (log) price of HC = constant* $-\ \alpha_0$	*1972 (level) price of HC*
Edu group 1	0	0.21859	0.21859	1.244321
Edu group 2	0.159	0.57754	0.41854	1.51974
Edu group 3	-1.164	-0.1005	1.0635	2.89649

Table 6: This table reports the 1972 normalised human capital prices computed by using constant terms estimated in the cross-sectional equation with the normalising (log) age intercepts.

series of human capital aggregates which are consistent with the technology parameters. For clarity we report in table (6) both the constant terms we have estimated for equations (20) and the normalising α_0^{edu} obtained from equations $(30-31)$. Their difference pins down a normalised (log) price for each human capital type in 1972 (the year for which we estimate the cross-sectional equation).

Using the price changes estimated in equations (21) we can obtain a time series of prices and human capital aggregates (efficiency weighted labor supplies) between 1968 and 1997. These series and their logs are presented in figure (8).

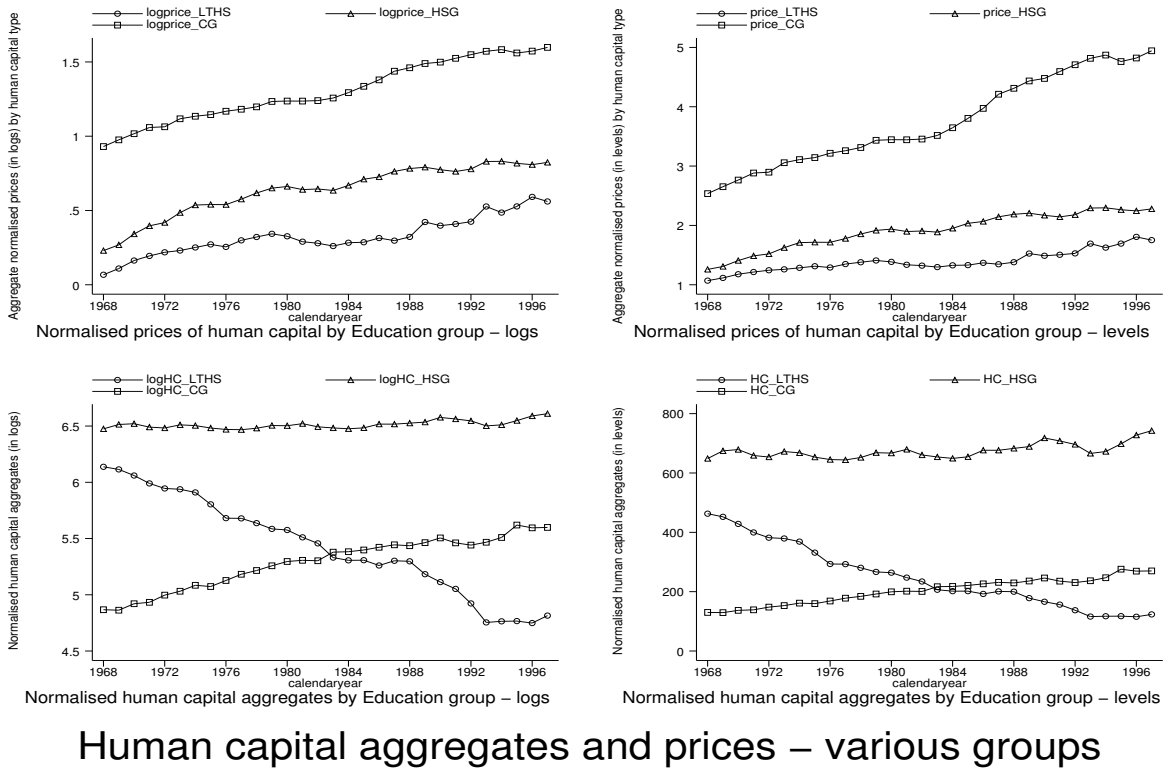

Figure 8: Human capital aggregates and associated prices over time, both logs and levels, based on preferred normalisation.

The price pattern reported in (8) is consistent with a pattern of increasing inequality. The time series of human capital stocks give an insight on the importance of selection in determining inequality, especially if we contrast them with the aggregate labor force and wage bills reported in figures $(3.3-3.3)$.

Despite a doubling of both the total number and wage bill of high school graduates, their human capital aggregate has been quite flat over the sample period, suggesting that for this group there has been a reduction in average per worker efficiency. A similar conclusion can be drawn for the college graduates, as their total number went up by almost four times over the sample period, whereas their human capital aggregate increased by roughly 70%. Big shifts in the distribution of people of different ability over educational outcomes have probably taken place over the sample period.

It is also worth noting that the pattern of HC aggregates is very similar to the labor shares' dynamics presented in figure (3.4) for the Cobb-Douglas case.

4 Calibrating the Benchmark Model

Not all the parameters in our model are estimated. The free parameters are chosen with the objective to build a numerical counterpart of our model which is able to reproduce selected features of the US economy.

Given the nature and timing of the choice faced by people wealth plays a pivotal role in determining equilibrium outcomes. The availability of assets and access to credit to smooth consumption is a crucial factor in both education and crime decisions. We set time-preference and borrowing limit parameters in order to obtain a benchmark with an appropriate wealth/income ratio and a share of asset-poor people in line with the observed share for the target year. The distribution of workers over education outcomes is equally important, because it determines the relative returns to the education investments. However, the aggregate education shares are not sufficient by themselves to pin down relative returns because the relative ability of workers is key in determining aggregate human capital inputs in the production function. Therefore we target not only the aggregate education shares in the target year, but also education shares by ability. The additional benefit of this calibration approach is that we are able to assess the composition effects of potential policies by looking at selection over ability as well as wealth.

Finally, the benchmark equilibrium of our model must be able to reproduce the aggregate (property) crime rate for the target. It is also necessary to restrict the marginal sensitivity of the aggregate crime rate with respect to the prison sentence to a value that is close to estimates from the empirical literature. This allows to measure the effects of alternative policies versus the case of a pure stiffening of sentences. We achieve this double objective by calibrating a utility cost of crime composed by 2 elements: a constant

purely conditional on committing crime and a quantity which is a function of the length of the sentence.

The remainder of this section describes our calibration approach in more detail.

Demographics. Each period represents one full year. An individual is born at age 16 and can choose whether to work or study. If an individual decides to study, she commits to be in school for two years until completion of High School. When 18 a High School graduate decides again whether to work or study for the next 3 years in order to become a College worker. In any case, agents can work only until age 65, which means that the full working life of a person who starts work at age 16 is 50 years, whereas it becomes 48 years for High School graduates and only 45 years for College graduates (who start working at age 22) . The age range in the model is the same as the age range we use in our PSID sample. The maximum possible age in the model is 95 and there is an age-related probability to die in each period that we take from the US life tables for 1989-1991.

Preferences. Agents have CRRA preferences and we choose the curvature of their utility to obtain a coefficient of relative risk aversion of 1.5. The discount factor β is chosen to produce a wealth income ratio equal to that for US households up t the 99% percentile. Wolff (2000) estimates the value of this ratio to be roughly 3.45 in 1983. The implied value of the discount factor is 0.967.

Unemployment shocks. Following Heathcote, Storesletten, and Violante (2003) we calibrate the required search period for an agent who experiences an unemployment shock to match the average duration of unemployment in the US economy. This is 13.5 weeks, which is roughly 26% of the full-time employed yearly work hours.We therefore set the labor supply of (temporarily) unemployed people to $\underline{l}$=0.74. The incidence of unemployment π_u (fraction of population experiencing an unemployment spell in a given year) is set to 17.5% and the model unemployment rate is $0.175 \times 0.26 = 4.55\%$ which is the US average for our sample period.[25]

Wealth distribution of the youngest. We assume that the wealth distribution among the youngest corresponds to the distribution of the accidental bequests in the economy. However, no agent is endowed with negative assets, so we censor the bequests' distribution at zero and appropriately modify the average bequest so that the total bequitted wealth is held constant.

Borrowing Limit. The exogenous borrowing limit $\bar{a}$ is calibrated to match the share of workers

[25] At the moment we do not differentiate employment risk by education, although an interesting extension would be to include education specific employment risk.

(all agents excluding students) with zero or negative wealth. Wolff (2000) provides an estimate of 15.5% for this share, which implies a borrowing limit of about 46% of average post-tax labor earnings.

Government. We use flat tax rates for both labor and capital income and, following Domeij and Heathcote (2003), we set $t_l = 0.27$ and $t_k = 0.4$. For simplicity, the pension is assumed to be a constant lump sum for all agents, regardless of their education and previous earnings. The replacement rate for the lump-sum is set to 16.4% of average post-tax labor earnings like in Heathcote, Storesletten, and Violante (2003).

Distribution of permanent characteristic (ability). We use the distribution of IQ test scores from the 1972 wave of the PSID to approximate the distribution of permanent characteristics (ability) over the population. For expositional simplicity we split the range of ability in 4 equal-size intervals and assign agents to such ability bins. The relative share of people in the four bins is different: only 1.7% of the total population are in the lowest ability group (bin 1) which contains people in the left tail of the distribution plotted in Figure 3.1. Just less than 6.6% of the total population is in bin 2, 48% in bin 3 and finally 44% in the highest ability group (bin 4).

Direct Cost of Education. The direct cost of college education is chosen to match the value of tuition costs as a proportion of average pre-tax earning. The National Center for Education Statistics provides several measures of tuition costs and we use our PSID sample for an estimate of average pre-tax earnings. Over the sample period the real college tuition costs have been been steadily growing, increasing from less than 5% to over 15% of our selected measure of earnings. We choose to set the college tuition costs to be 10% of average *post-tax* earnings. Given the labor tax rate in our model, this is equivalent to a college tuition cost roughly equal to 8% of average *pre-tax* earnings.

For the value of High School direct costs we have set them to be just 1% of average post-tax earnings, in order to account for expenses incurred for studying equipment and other practicalities. There does not seem to be not much information on such costs.

Education Enrolment Rates. Eeducation rates are matched both in the aggregate and by ability groups. The distinction is important because the same aggregate shares are consistent with many different distributions of ability over education and, therefore, many different relative marginal returns between different types of labor. Moreover, the policy experiments are likely to alter the distribution of

Enrolment rates by ability bin

	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	0.75	0.5	0.28	0.12	0.24
Edu 2	0.25	0.43	0.57	0.53	0.54
Edu 3	0.00	0.05	0.15	0.35	0.22

Table 7: Shares of workers in different education groups by ability (IQ test) bin. Based on our sample of workers from the 1972 wave of the PSID.

Grossed-up enrolment rates by ability bin

	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	0.81	0.56	0.30	0.13	0.25
Edu 2	0.19	0.40	0.58	0.61	0.58
Edu 3	0.00	0.04	0.12	0.26	0.17

Table 8: Grossed-up shares of workers in different education groups by ability (IQ test) bin. The values are grossed-up in equal proportions to replicate the aggregate education shares observed for workers in the 1980 CPS March supplement.

ability in each education group and it is useful that the benchmark can reproduce the distribution of ability types over education outcomes. In order to approximate such distribution we use the 1972 wave of the PSID which provides data on educational attainment of agents as well as their score in an IQ test. We assign people to 4 different ability bins, with bin 1 comprising those with the lowest IQ scores and bin 4 those with the highest. The education shares for each ability bin and the ensuing aggregate fractions are reported in table (7).

However, the aggregate education shares based on the 1972 wave of the PSID do not represent the true shares of aggregate enrolment in the US economy in our sample period[26] which are reported in table (B). In order to reproduce the aggregate education distribution in the economy we gross-up the 1972 rates so that their aggregation gives back the aggregate enrolment rates for the US economy in 1980[27]. The values of the grossed-up education shares by ability are reported in table (8).

We use ability-specific quasi-linear utility terms $\psi_j(\theta)$ to shift the value of education for different ability bins and match the education shares.

Aggregate Crime Rate and Elasticity of Aggregate Crime Rate to the Expected Prison Sentence. The aggregate property crime rate for the US in 1980 was 5.6%.[28] Furthermore, the evidence linking increased punishment to aggregate crime rate indicates that the elasticity of property crime with

[26]One reason for this problem is attrition which can unequally affect people with different education in the PSID, altering the aggregate education shares. Moreover, our sampling procedure is likely to exclude people with low attachment to the labor market.

[27]We use 1980 for the aggregate enrolment rates because the education shares in that year lie very close to the sample averages for the period 1967-2001. The average fraction of workers with no High School degree over the sample period was 0.232. The fraction of High School graduates was 0.575 and the College graduate share was 0.193.

[28]The crime rate is a victimization rate and represents a per capita measure of property crime in the US. The data are from the Uniform Crime Report and are taken from the Sourcebook of Criminal Justice Statistics, Bureau of Justice Statistics. An alternative source of data regarding crime victimisation is the National Crime Victimisation Survey, based on self-reporting by victims. This study suggests a larger incidence of crime vis-a-vis the UCR data.

respect to expected punishment ranges from -0.1 to -0.4, see Donohue and Siegelman, 1998. We target an elasticity of -0.2, following Levitt (2004) who picks this value to account for the increase in incarceration rates over the 1990s.

We jointly choose the utility in jail $\bar{u}$ and the quasi-linear utility term associated to committing crime so that we match the aggregate crime rate and its elasticity to expected punishment. The average expenditure per convict in the model is equal to m. According to Maguire and Pastore (1995) the average expenditure per convict in the US was roughly \$ 20,000 in 1992 and went up to \$ 26,000 by 1999. These per-prisoner-costs are roughly 53% of average pre-tax labour earnings from our PSID sample. Donohue and Siegelman (2004) suggests an even higher cost per prisoner of \$ 36,000 for 1993 that would be 90% of the average pre-tax earnings in our PSID sample. We choose to set average per prisoner costs to 0.53 of average earnings and set m to match this value.

The value of the parameters calibrated in the benchmark are reported in table (35) with the exception of the utility quasilinear terms associated with the education decisions which are reported in table (B). Both tables are in the appendix.

5 Numerical Simulations

This section describes the benchmark economy and presents the results of our policy experiments. We start by describing the main features of our benchmark economy in some detail. All models results are reported in model units.

5.1 The Benchmark Economy

Education distribution. In our benchmark economy both the aggregate and ability-specific distribution of people over educational outcomes reproduce the shares in table (8). At the time of the high school choice wealth seems to matter only for people in the two highest ability bins. Given the relatively small differential in wages between high school graduates and dropouts (roughly 8% in our benchmark simulation, see table 10), people in the lowest ability bins are roughly indefferent between high school and working. Only for the highest ability group there is a preference for high school[29] which introduces selection based on wealth. In these bins, agents who opt to continue schooling at ages 16 are richer

[29]This preference is explained by the fact that ability multiplies market prices in this model, and by the fact that many of the high ability people are bound to progress to college.

Average wealth of workers and students at age 16 and at age 18					
		age 16		age 18	
		workers	*students*	*workers*	*students*
average wealth	ability bin 1	8.18	7.77	5.09	4.82
	ability bin 2	7.66	9.77	7.20	5.28
	ability bin 3	1.43	11.48	9.17	6.44
	ability bin 4	0.31	9.69	6.68	7.00

Table 9: Average assets holding of people who choose to work vis-a-vis people who choose to study at ages 16 and 18. All values are in model units.

overall than those who opt for work. As shown in table 9, in ability bin 3 the average wealth of people who progress to high school is 11.48 model units compared to just over 1.4 for the people who choose to work (in this bin high school dropouts account for roughly 30% of the population, see table 8). The difference is even starker in ability group 4, where only 13% of people are high school dropouts. In this group the average wealth of students in 9.7 compared to a very low 0.31 for the dropouts. Given that we are conditioning on being part of a very high ability group, it is clear that selection is working mostly through the initial wealth endowment.

Admittedly this result depends on the assumption we are making regarding the borrowing limit of agents: we are assuming that people can borrow up to 46% of average post-tax labor earnings. Finally, many of the people who decide to go to high school in the high ability bins are likely to continue onto college, where tuition costs are higher and the length of the period to be funded longer.

In contrast to the high school decision, no large differences in wealth are present between people who decide to work at age 18 and people who decide to progress to a college education. The benchmark suggests that among college goers selection is based mostly on ability and wealth plays a smaller role[30].

This results suggest that selection based on wealth takes place at an earlier stage in life: education decisions are a sequential process, and by the time of college only for a very small fraction of agents the decision will depend upon their wealth.

We also find that there is almost no difference in the average ability of college and high school graduates, which we mostly attribute to the fact that the gradient of ability in college wages is zero in this model.[31]

Average earnings and Income by Education. The model is able to replicate some stylised facts about inequality in earnings and income. Table (10) reports post-tax wage and income, together

[30]We also measure what share of potential college goers are borrowing constrained at age 18. The only ability bin that contains borrowing constrained potential college goers is the highest, and even here only 12% of agents are borrowing constrained. This corresponds to just above 5% of the total of potential college goers.

[31]See table (1) in the empirical section for an explanation of the zero gradient of IQ on college wages.

Inequality summary statistics for the benchmark economy						
	Education					
	HS dropouts		HS Graduates		College Graduates	
	Crime	No Crime	Crime	No Crime	Crime	No Crime
Average ability	0.948	1.023	1.127	1.160	1.111	1.137
Average wage	1.19	1.72	1.28	1.85	1.37	2.71
Average ability (by edu only)	1.011		1.159		1.137	
Average wage (by edu only)	1.63		1.84		2.70	
Average income (by edu only)	1.30		2.03		3.07	

Table 10: Summary statistics for different education groups, by criminal status.

Self-reported criminal participation rates by education status in NLSY (males ages 20-23,1980) Source: table 1, p.825, Lochner (2004)		
	Any income from crime	Property crime
years of schooling		
Less than 10 years	0.297 (0.035)	0.129 (0.026)
10-11 years	0.337 (0.029)	0.218 (0.026)
12 years	0.244 (0.017)	0.118 (0.013)
more than 12 years	0.174 (0.015)	0.160 (0.015)

Table 11: Property crime includes thefts of at least $ 50 or shoplifting. Standard errors in parenthesis. The data are taken from table 1, page 825, of Lochner (2004).

with average ability, by criminal and education status. In particular, the relative post-tax income by education is in line with the long term differences observed between 1967-1996, suggesting that the model is doing a good job in capturing inequality in both labor and capital income.

Crime statistics and costs. Empirical evidence suggests that the largest share of property crime is committed by young, uneducated people. Lochner (2004) provides a large body of evidence documenting a strong correlation between young age, low education and crime. In table (11) we report some data on education and crime based on the NLSY, provided by Lochner (2004).

On the basis of this evidence it appears that completing 12 years of schooling does make a large difference in the propensity to commit crime. If we take the average crime rate between ages 20 and 23 produced by our benchmark model for different education groups, we have that among high school dropouts the crime rate is 0.35, among high school finishers it is 0.07 and among college graduates it is 0.10. It seems that our model is slightly overestimating the share of (property) criminals who are high school dropouts compared to NLSY data and underestimating the share of criminals with at least a high school degree.[32] This is due to the nature of the crime decision, which has no long term consequences for those apprehended. Introducing some form of stigma, like in Imrohoroglu, Merlo, and Rupert (2004) would probably reduce this discrepancy. It must be also noted that our measure of crime rate is in fact

[32]NLSY data might somehow underestimate the true rates as individual are self-reporting their crime activity.

Crime rates by ability bin and education group - benchmark			
		Education group	
ability bin	*High school dropouts*	*High school graduates*	*College graduates*
bin 1 (lowest)	0.26	0.17	0.024
bin 2	0.19	0.033	0.016
bin 3	0.17	0.009	0.009
bin 4 (highest)	0.11	0.025	0.009
aggregate	**0.165**	**0.019**	**0.01**

Table 12: Simulated crime rates by ability bins and education groups in the benchmark economy. The aggregate refers to the education-specific average of different abilities weighted by the relative size of each ability bin.

a victimization rate based on the whole population in a certain age group, whereas the NLSY data's population is just the sample population in a given year.

The distribution of crime rates by ability bins and education group in our model shows large selection effects for the crime choice. Table (12) summarizes such crime rates in our benchmark. Given ability, crime rates drop dramatically as the labour market opportunity cost increases. This effect has been already documented by Machin and Meghir (2004).

Increasing opportunity cost seems to work mostly through education, although there are sizeable effects in ability, especially in the group of high school finishers.

The aggregate crime rate in the benchmark economy is 5.6%. The aggregate prison expenditure is equal to 0.43, that is 0.3% of aggregate output and 1% of total tax revenues.

5.2 The Effects of Increased Punishment

We have calibrated utility in jail in order to obtain a (local) elasticity of aggregate crime rate with respect to expected prison term of roughly -0.2 (see Levitt (2004)). In order to assess the effects of increasing the expected prison term in our model, we run some experiments in which we increase the prison term.

In the first of such experiments, we increase the expected prison sentence by 0.1 units of a year, which is equivalent to 1.2 months. This change increases the expected sentence for an apprehended criminal from 12.6 months in jail to 13.8 months in jail. It is worth pointing out that in our model apprehension corresponds to incarceration. However, based on the Sourcebook of Crime Statistics, only 66% percent of cleared property crime cases reaching court end up in positive jail sentences. Therefore, expected sentences of 12.6 and 13.2 months correspond to average dispensed sentences of, respectively, 19.1 and 20.9 months.

Crime rates by ability bin and education group - jail term 13.2 months (G.E.)			
		Education group	
ability bin	*High school dropouts*	*High school graduates*	*College graduates*
bin 1 (lowest)	-0.03	0	-0.02
bin 2	-0.02	0	0
bin 3	-0.01	0	0
bin 4 (highest)	0	0	0
aggregate	-0.01	**0**	**0**

Table 13: Differences with respect to benchmark (absolute changes) in simulated crime rates by ability bins and education groups for economy with expected prison term of 13.2 months - G.E.

Enrolment rates by ability bin - jail term 13.2 months (G.E.)

	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	-0.06	-0.01	0	0	0
Edu 2	0.05	0.02	0.01	-0.01	0
Edu 3	0.01	-0.01	-0.01	0.01	0

Table 14: Differences with respect to benchmark (absolute changes) in shares of workers in different education groups by ability (IQ test) bin, given a 1.2 months increase in prison term

Increasing the expected jail term to 13.2 months generates, in general equilibrium, a drop of the aggregate crime rate to 5.3%. The effect of the higher punishment, as summarised in table (13), is to reduce the crime rate among high school dropouts, especially in lower ability groups.

The increased punishment also generates a change in the ability specific distribution of education. Table 14 reports the new values. In the lowest ability group the number of high school dropouts goes down from 81% in the benchmark to 75% in the new equilibrium. Also in the second lowest ability group high school dropouts decrease from 56% to 55%. We attribute this change to a stronger disincentive to use crime as a consumption smoothing device. Since low skill people have a higher propensity to engage in crime for given educational attainments, this improves the ability composition of the pool of high school dropouts further reducing the crime rate for this education group. The aggregate education distribution is roughly unaffected given the low number of people in the first two ability bins.

No large effects in other dimensions take place in the new equilibrium. The aggregate prison expenditure goes up to 0.45, and it goes up marginally as a share of total tax revenue to 1.1% . Additional costs are paid through increases in the labor tax rate which is almost unchanged at 27.03%. Per-capita output, after subtracting the prison costs, is almost unchanged with respect to the benchmark (up by 0.02%). The effects of the change in prison term length are very similar in partial equilibrium, meaning that no significant price effect is induced by this policy change.

We also experiment with an even longer prison term, pushing up the expected sentence by a further 1.2 months to 15 months. The aggregate crime rate goes down only marginally in this case, to 5.27%.

In the second ability bin there is a larger drop in high school dropouts' share, which goes down to 51% of the total. However the aggregate prison costs increase to 0.491 or 1.2% of total tax revenues and, as a consequence, the labor tax rate increases to 27.09%.

5.3 Subsidizing High School Completion

The second experiment we carry out involves subsidizing high school completion. A very well known experiment - the Quantum Opportunities Program - was carried out on a small scale along similar lines by the Department of Labor and the Ford Foundation in two waves: a first one between 1989 and 1993 and a later one between 1995 and 2001. The program was targeted at adolescents from families receiving public assistance. The experiment, appropriately randomized, offered learning support and cash incentives, from grade nine through to high school graduation, to students in the treatment group. It involved a "salary" starting at $1 and rising to $1.33 per each hour of "activity" the student attended up to a ceiling plus a bonus of $100 for each 100 program hours for completing activities[33]. An amount equal to the earned stipend was also deposited in an accrual account and paid to the enrollee conditionally upon completion of her high school degree. The total cost of the program was $3130 per student per year, of which $2150 represented the direct payment to the student and the remaining amount the cost of the resources (teaching support and equipment) the student had access to.

Hahn, Leavitt, and Aaron (1994) report that the program reduced the crime rate in the year after the end of the program by roughly 50% (from 13% to 6%) among participants relative to the control group. As discussed in Donohue and Siegelman (2004), there are three main reasons why these numbers must be taken with care. First, the data refer to the subjects self-reporting about being in trouble with the police. Secondly, the difference in self-reported crime rates between the two groups was only significant at 12% according to Donohue and Siegelman's (2004) calculations. Third, the significance decreases further if data from the unsuccessful Milwaukee trial, which the analysts dropped from their calculations, were included.

The experiment in this section differs from the Quantum Opportunity Program as it is not targeted at students from a disadvantaged background, but consists in giving *all* students attending high school a subsidy equal to 7.6 per cent of average labour earnings, which corresponds to the ratio between $ 3130 and average labour earnings in the data from 1995, one of the central years in which the program

[33]The maximum number of program hours was 750 divided between 250 hours of academic support, 250 hours of cultural and developmental activities and 250 hours of service activities, such as community service projects.

Enrolment rates by ability bin - non means-tested (non M.-T.) HS subsidy (P.E.)					
	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	-.13	-.23	-.15	-.05	-.11
Edu 2	.12	.23	.15	.03	.10
Edu 3	.01	0	0	.02	.01

Table 15: 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).

Crime rates by ability bin and education group - non M.-T. HS subsidy in P.E.			
		Education group	
ability bin	*High school dropouts*	*High school graduates*	*College graduates*
bin 1 (lowest)	.01	.01	-.01
bin 2	.02	.04	-.01
bin 3	.04	.02	0
bin 4 (highest)	.01	.01	0
aggregate	**.03**	**.02**	**0**

Table 16: Differences with respect to benchmark (absolute changes) in simulated crime rates by ability bins and education groups in the P.E. non means-tested high school tuition subsidy experiment. The aggregate refers to the education-specific average of different abilities weighted by the relative size of each ability bin.

was run. Since students are likely to have benefited from the teaching support as well as the transfer, we choose to use the total cost of the program which reflects the actual market value of the resources enjoyed by a student.

The partial equilibrium effects of this simple tuition subsidy are limited in terms of crime reduction: the aggregate crime rate goes down to 5.38% This is despite a huge shift in the distribution over education conditional on ability. The effects in terms of high school completion are extremely large, as reported in table 15.

The total number of high school dropouts in the economy drops by more than 40%. The increase in total tax revenues due to the job reallocation more than compensates the costs of the program. Aggregate prison costs plus the costs of the tuition programme are just 1% of total tax revenues. Per-capita output, after subtracting jail costs, is 1.2% higher than in the benchmark: as more people have access to labor markets with higher returns total output grows substantially. As shown in table 16, the positive effect in terms of crime associated to moving people away from the pool of high school droputs is counterbalanced by a sharp rise in crime rates among the remaining high school dropouts. This increase in crime rate among dropouts is largely due to bigger inequality, since higher earnings for those who move into higher education are not matched by higher earnings for those left behind.

In table 17 we report percentage changes (with respect to the benchmark) in between-group differ-

Changes in inequality (%). Non means-tested H.S. tuition experiment - P.E.		
	HS graduates *vs* HS dropouts	College graduates *vs* HS graduates
Average ability difference	-6.1	0
Labor price difference	0	0
Average wage difference	-4.8	1.2
Average income difference	54.8	1.0

Table 17: Change with respect to benchmark (percentage) in various between-group differences. Labor prices are marginal returns per efficiency unit of labor, obtained from the production technology. Experiment: non means-tested HS tuition subsidy - P.E.

ences in average wage, ability, marginal returns[34] and post-tax income. Given that mostly wealth-poor agents remain in the drop-outs group, there is an increase of more than 50% in the income differential between high school graduates and dropouts. This additional inequality explains the substantial increase in crime rates among dropouts. The increase in crime rates among high school graduates follows from the fact that poorer and less able people now represent a larger share of this education group.

In partial equilibrium the increase in inequality seems to dampen the effectiveness of the HS tuition policy.

In the general equilibrium case things are rather different. The drop in crime rate is quite large with respect to the benchmark, as well as the experiments with higher jail terms. Table 18 reports the crime rates by ability and education for the G.E. high school subsidy experiment.

What drives the result is a significant composition effect. As it is clear from Table 19 the proportion of ability 1 people among high school dropouts falls by 8 percentage points (against 6 in the prison term experiment) while the proportion of ability 2 people increases by 4 percentage points (it *falls* by 5 percentage points in the prison term case). Since the average worker in ability bin 2 has half the propensity to engage in crime than the average ability 1 worker, the crime rate falls more than in the case of an increase in the prison term. This improvement in the ability mix among high school dropouts produces an equilibrium aggregate crime rate of 5.1%, with a drop twice as large as the ones induced by the increased punishment and by the same tuition subsidy in partial equilibrium. Relative both to the prison term experiment and to the partial equilibrium case, this favorable composition effect is driven by changes in the relative price of skill between high school dropouts and high school graduates. Such relative price is basically unchanged in the prison term experiment. However, the labor price differences substantially shrink in the case of the tuition subsidy. This increases the share of people within ability group 2 who are high school dropout, because higher ability people have a higher opportunity cost of

[34]The differences in marginal returns to labor are unchanged by construction in this P.E. experiment.

Crime rates by ability bin and education group - HS subsidy in G.E.			
		Education group	
ability bin	*High school dropouts*	*High school graduates*	*College graduates*
bin 1 (lowest)	.01	0	-.02
bin 2	-.02	0	0
bin 3	-.01	0	0
bin 4 (highest)	0	0	0
aggregate	**-.02**	**0**	**0**

Table 18: Differences with respect to benchmark (absolute changes) in simulated crime rates by ability bins and education groups in the G.E. high school tuition subsidy experiment. The aggregate refers to the education-specific average of different abilities weighted by the relative size of each ability bin.

Enrolment rates by ability bin - HS subsidy (G.E.)

	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	-.08	.04	0	-.02	0
Edu 2	.04	-.04	0	.01	-.01
Edu 3	.04	0	0	.01	.01

Table 19: 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 (general equilibrium).

attending high school relative to those in ability group 1. Education and skills are substitutes in reducing the crime rate and the equilibrium change in market prices shifts towards high school the people with the higher propensity to engage in crime.

Finally, in table 20, we document a small decrease in labor income inequality and a larger decrease in total income inequality that reinforces the drop in disaggregated crime rates.

The drop in income inequality is due to the adjustment in prices that was barred in partial equilibrium. Inequality in total income between HS graduates and dropouts is roughly one third smaller in G.E. than it was in partial equilibrium.

In this equilibrium the aggregate prison costs in model units are 0.40, or 0.94% of total tax revenues. The high school subsidy program costs roughly 0.5% of total tax revenues, which gives a total cost of roughly 1.44% of total tax revenues. However, in the new equilibrium total tax revenues are 0.4% larger than in the benchmark, despite the proportional labor tax rate increases by the same amount (to

Changes in inequality (%). Non means-tested H.S. tuition experiment - G.E.		
	HS graduates *vs* HS dropouts	College graduates *vs* HS graduates
Average ability difference	12.2	31.8
Labor price difference	-16.7	-1.4
Average wage difference	-4.8	-1.2
Average income difference	-5.5	-2.9

Table 20: Change with respect to benchmark (percentage) in various between-group differences. Labor prices are marginal returns are per efficiency unit of labor, obtained from the production technology. Experiment: non means-tested HS tuition subsidy - G.E.

Enrolment rates by ability bin - means-tested HS subsidy (P.E.)

	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	-.06	-.14	-.15	-.04	-.1
Edu 2	.05	.13	.15	.03	.09
Edu 3	.01	0	0	.01	.01

Table 21: Difference in shares of workers (absolute changes relative to benchmark) in different education groups by ability (IQ test) bin, given a means-tested high school subsidy intervention (partial equilibrium).

27.03%) as in the case of longer prison term. Per-capita output, after subtracting jail costs, goes up by 0.5%. The increase in efficiency and revenues makes the policy effectively self-financing. Yet, for the same change in the tax rate, the effect in terms of crime reduction is roughly double as for the increase in prison term.

5.4 Restricting the Target Population: a Means-Tested High School Subsidy

Another step towards designing an experiment comparable to the Quantum Opportunity Programme is to restrict the target population. The QOP was intended to help students from disadvantaged backgrounds: in this sense the population eligible for financial support was not the universe of all potential students as in the previous experiment. An across-the-board subsidy (i.e. not conditional on available resources) corresponds to a giveaway to many inframarginal individuals who would attend high school in any case, and its per-dollar effectiveness is likely to be smaller than a targeted intervention.

It is interesting to replicate the previous experiment with a restriction on eligibility. We do this by making the subsidy available only to agents with an inital assets endowment below the 35th percentile of the initial assets distribution. We don't change the size of the transfer, which is still 7.6 per cent of average labour earnings.

The P.E. result in terms of aggregate crime rate reduction is smaller than for the case of the subsidy across the board: the aggregate crime rate goes down to 5.43%. This follows from the smaller number of agents switching from the HS dropouts pool to the high school finishers group. Table (21) reports the changes (with respect to the benchmark) in education shares by ability.

The most noticeable difference with respect to the unrestricted subsidy case in partial equilibrium is the drop in criminal activity for the high school dropouts within ability bin 2 (table 22). This curious fact can be rationalised by looking at table (23), reporting the share of eligible people taking up the subsidy by ability group. The take-up is increasing in ability (just as enrolment in HS is increasing in ability). However, for given wealth, the number of marginal individuals who shift education group because of

Crime rates by ability bin and education group - means-tested H.S. subsidy in P.E.			
		Education group	
ability bin	*High school dropouts*	*High school graduates*	*College graduates*
bin 1 (lowest)	.01	.02	-.01
bin 2	-.03	.06	-.01
bin 3	.04	.02	0
bin 4 (highest)	.01	.01	0
aggregate	**.02**	**.02**	**0**

Table 22: Differences with respect to benchmark (absolute changes) in simulated crime rates by ability bins and education groups for the P.E. means-tested high school tuition subsidy experiment. The aggregate refers to the education-specific average of different abilities weighted by the relative size of each ability bin.

Share of eligible population taking-up the subsidy, by ability bin				
Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate*
0	0.44	0.58	0.75	0.63

Table 23: Share of eligible agents who take-up subsidy and complete high school. Means-tested HS subsidy experiment in P.E.

the subsidy is larger in ability bin 2 than in ability 3 or 4; this is because there are many inframarginal individuals in bins 3 and 4 who would go to high school in any case. This is confirmed by comparing the enrolment rates in this experiment with the corresponding partial equilibrium experiment with unrestricted subsidy: among those ability groups with positive take-up shares, only people in ability group 2 experience a significant change in enrolment. This explains the drop in crime rate for that group[35]. This means-tested policy is having a strong effect on a specific group of people (in ability bin 2) whose ability (and wealth) are high enough to have positive returns from switching education, and low enough to find such switch unattractive without a subsidy. With respect to the non means-tested subsidy, many people whose assets are above the means-testing threshold are likely to remain in the HS dropout pool (just as they were in the benchark), pulling down the crime rate in that group. On the other hand, people who were poorer and more likely to commit crime seized the opportunity to become HS graduates thanks to the policy.

The average initial wealth and ability of eligible people who took advantage of the programme are significantly higher than for those who turned down the subsidy. Table (24) summarises the main differences between those eligible agents who took advantage of the programme and those who did not, and compares both to the group of people who did not qualify for the programme because of higher wealth. The results confirm that only people with much lower wealth *and* ability did not get any advantage from the programme. Moreover, the takers experienced much lower crime involvement over

[35]If there was a positive take-up in ability group 1 we would observe a similar phenomenon. However the results show that in ability group 1 there are no marginal individuals whose behaviour changes because of the policy.

Eligible takers, eligible non takers and non–eligible: a comparison. MT HS subsidy, P.E.

	eligible		non eligible
	take-up	*no take-up*	
Average initial wealth	2.66	0.39	12.11
Ability	1.157	1.052	1.119
Crime rate over life cycle	0.09	0.21	0.01

Table 24: Some statistics comparing eligible people who took advantage of the programme to eligible people who did not. We also report the same statistics for people who did not wualify for the programme because of excessive wealth. We report averagel wealth endowment at time of choice (age 16), average ability and average crime rate over the life cycle. Experiment: means-tested HS subsidy P.E.

Enrolment rates by ability bin - means-tested HS subsidy (G.E.)

	Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate rates*
Edu 1	.19	.11	0	-.03	0
Edu 2	-.19	-.10	0	.02	-.01
Edu 3	0	-.01	0	.01	.01

Table 25: Differences with respect to benchmark (absolute changes) in shares of workers in different education groups by ability (IQ test) bin, given a means-tested high school subsidy intervention (general equilibrium).

their life cycle, although this result can be misleasing insofar we are not controlling for self-selection in the programme. Finally, the group of non eligible agents is unsurprisingly much richer on average and its average ability lies between the ones of the other 2 categories. Not much in terms of crime reduction is lost by reducing the scope of the intervention, since the non eligible agents have very low crime rates over their life cycle.

This P.E. experiment produces levels of wage and income inequality which are comparable to the levels obtained in P.E. within the unrestricted subsidy experiment. The aggregate prison expenditure is 0.42 (1%) of total tax revenues, whereas the aggregate transfer expenditure is 0.092 (0.02%) of total tax revenues. The labour tax rate consistent with budget balance is 26.98%, just below the benchmark level. Per-capita output, net of jail costs, goes up by 0.27% with respect to the benchmark.

When we turn our attention to the G.E. means-tested experiment, things are different. The aggregate crime rate goes down much more than in P.E., to 5.2%. Strong composition effects are again present in the G.E. case. The policy is taken up by less people than in P.E. (40% vs 63%) and only the two highest bin groups seem to have a significantly positive take up share. Nonetheless, crime rates by ability and education drop substantially with respect to the benchmark as well as the P.E. counterpart. Tables $(25-28)$ summarise the results. The mechanisms at work are similar to the case of the non means-tested subsidy.

Wage and income inequality are similar to the G.E. results for the across-the-board subsidy. The wealth difference between eligible takers and non-takers are smaller than in the P.E. case, the ability

Crime rates by ability bin and education group - HS subsidy in G.E.			
	Education group		
ability bin	*High school dropouts*	*High school graduates*	*College graduates*
bin 1 (lowest)	-.02	-.17	-.20
bin 2	-.03	0	-.01
bin 3	-.02	0	0
bin 4 (highest)	0	.01	0
aggregate	**-.02**	**0**	**0**

Table 26: Difference in simulated crime rates (absolute changes relative to benchmark) by ability bins and education groups in the G.E. means-tested high school tuition subsidy experiment. The aggregate refers to the education-specific average of different abilities weighted by the relative size of each ability bin.

Share of eligible population taking-up the subsidy, by ability bin				
Bin 1	*Bin 2*	*Bin 3*	*Bin 4*	*Aggregate*
0	0.005	0.175	0.718	0.397

Table 27: Share of eligible agents who take-up subsidy and complete high school. Means-tested HS subsidy experiment in G.E.

difference is larger: price effects adjust in such a way that only high ability people are marginal with respect to the policy. This happens because marginal returns to different education types are closer than in P.E. and make a difference for high ability people only. Lower inequality reduces the incentive to crime for low education and low ability people.

The aggregate prison costs are just below 0.41, or 1% of total tax revenues. The cost of the transfer programme is 0.06 which is 0.14% of total tax revenues, and the labour tax rate consistent with balanced budget is 26.99%. Per-capita output, net of jail costs, is 0.1% higher than in the benchmark. The G.E. effects of this means-tested intervention are more cost effective than an unrestricted transfer policy, although not as effective in terms of aggregate crime reduction because the unrestricted transfer policy generated an aggregate crime rate of 5.1%. This result is due to the fact that in the means-tested case those relatively wealthy individuals in ability bin 1 and 2 who would have gone to high school in case of subsidy are not receiving any and the price differential is not large enough to make them switch. This fact more than compensate an even larger decrease in crime disaggregated crime rates.

Eligible takers, eligible non takers and non–eligible: a comparison. MT HS subsidy, G.E.			
	eligible		non eligible
	take-up	*no take-up*	
Average initial wealth	2.88	1.20	12.08
Ability	1.24	1.04	1.12
Crime rate over life cycle	0.06	0.17	0.02

Table 28: Some statistics comparing eligible people who took advantage of the programme to eligible people who did not. We also report the same statistics for people who did not wualify for the programme because of excessive wealth. We report averagel wealth endowment at time of choice (age 16), average ability and average crime rate over the life cycle. Experiment: means-tested HS subsidy G.E.

Treatment and control group: a comparison. Means-Tested HS subsidy			
	General equilibrium case		
	Randomised in		Randomised out
	take-up	*no take-up*	
Crime rate over life cycle	0.07	0.19	0.14
	Partial equilibrium case		
	Randomised in		Randomised out
	take-up	*no take-up*	
Crime rate over life cycle	0.09	0.21	0.15

Table 29: Effectiveness of means-tested tuition subsidy in reducing life cycle crime. We compare people randomised in and out of the experiment. Results are provided for both G.E. and P.E. experiments.

5.5 Introducing a Control Group: A 50/50 randomisation of the HS Subsidy

The most interesting piece of information regarding a subsidy experiment is how effective it is in reducing crime among those who receive it. In order to make a statement regarding such change, one should be able to compare the crime rates of two groups completely identical in every respect but the subsidy.

In order to obtain such information we have repeated the means-tested HS subsidy experiment described above with the simple variant that, among the eligible individuals, only a randomly chosen 50% would receive the subsidy (with the remaining 50% not getting anything).

For these experiments we present only result comparing the average crime rates (over their life cycle) for eligible people randomised in or out. Here we don't report information on the effects of the policies because they have been documented in previous sections. What matters in these experiments is the difference in crime rates between treatment and control group. We have run these experiment both in G.E. and P.E. and report the results for both in table (29).

The subsidy reduces life cycle crime rates by roughly half when we compare people who are randomised in and take up the subsidy with people who are randomised out. These numbers are very much in line with crime rates for the treatment and control group one year after the end of the Quantum Opportunities Programme experiment that we have reported at the beginning of Section 5.3. Though care must be taken due to the different time horizon over which they are calculated, their similarity is remarkable. It is worth noting that the average crime rate for people who are randomised out includes observations relative to people who would not take up the programme if it was offered to them. These results suggest that a targeted subsidy policy can reduce crime rates in the target population by more than half, with the benefit being spread over a long time horizon corresponding to the life cycle of the treated.

6 Conclusions

In this paper we have asked if a policy affecting the education decisions of relatively poorer and less able people can be more effective, in terms of costs and results, in reducing (property) crime rates than policies based on harsher punishment alone. We have developed and estimated a structural overlapping generations, life cycle model with optimal consumption, education and crime decisions. Given the complexity of the model, we have solved it numerically and have simulated the outcome of two alternative sets of policies - increases in prison sentences and subsidies for high school completion. Our findings suggest that subsidizing high school graduation is cost effective and preferable to policies based on harsher punishment. We have found that the effect of a subsidy depend on the size of the intervention, i.e. whether the programme is large enough to generate changes in prices, at least locally. Our results indicate that, relative to partial equilibrium, price changes induce an improvement in the ability composition of the high school dropout pool and lower income inequality across education groups. The two effects reinforce each other in reducing the crime rate. We have shown that subsidies targeted at poorer students can be nearly as effective in terms of crime reduction as unconditional subsidies, at significantly lower cost. Finally, controlling for unobserved heterogeneity through a randomisation of the policy intervention, we have found that means-tested subsidies towards high school completion reduce the average crime rate over the life cycle of a target individual by almost one half. The framework can be easily extended to allow for differential employment risk by education and to study the effect of other interventions such as wage subsidies, unemployment benefits, income tax credits and other redistributive policies. This is ongoing research.

References

Abraham, A. (2001): "Wage Inequality and Education Policy with Skill-biased Technological Change in OG Setting," Ph.D. thesis, Universitat Pompeu Fabra.

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

Cozzi, M. (2004): "Black-White Labour Market Conditions and property Crimes in the US: a Quantitative Analysis," unpublished manuscript, University College London and University of Pennsylvania.

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

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.

Freeman, R. (1999): "The Economics of Crime," in *Handbook of Labor Economics Vol. 3C*, ed. by O. Hashenfelter, and D. Card, chap. Chapter 52. North Holland, Amsterdam, The Netherlands.

Gallipoli, G. (2005): "On Non-Convexities in a Life-Cycle Model," Unpublished Manuscript, University College London.

Gallipoli, G., C. Meghir, and G. Violante (2004): "Human Capital Accumulation, Education Policy and Wage Dispersion," Mimeo, University College London.

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(4), 756–91.

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 (2003): "The Macroeconomic Implications of Rising Wage Inequality in the US," Mimeo.

Heckman, J., L. Lochner, and C. Taber (1998): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," NBER Working Paper, 6384.

Hill, B. (1975): "A Simple General Approach to Approximate the Tail of a Distribution," *Annals of statistics*, 3, 1163–1174.

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.

Lee, D. (2001): "An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice," Ph.D. thesis, University of Pennsylvania.

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

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

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

——— (2004): "Understanding why crime fell in the 1990s," *Journal of Economic Perspectives*, 18(1), 163–190.

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

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

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

Maguire, K., and A. L. Pastore (1995): "Sourcebook of criminal justice statistics, 1994," Discussion paper, U.S. Department of Justice, Bureau of Justice Statistics.

Merlo, A. (2001): "The Research Agenda: Dynamic Models of Crime and Punishment," *Economic Dynamics Newsletter, Supplement to the Review of Economic Dynamics*, 2, http://www.economicdynamics.org/News41.htm.

Polivka, A. E. (2000): "Using Earnings Data from the Monthly Current Population Survey," mimeo, Bureau of Labor Statistics.

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.

Stokey, N. L., R. E. Lucas, and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics.* Harvard University Press.

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

West, S. A. (1985): "Estimation of the Mean from Censored Income Data," mimeo.

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

A PSID Data

The Panel Study of Income Dynamics provides information on a variety of dimensions. Since the beginning it was decided that those eligible for the 1969 and following waves of interviewing would include only persons present in the prior year, including those who moved out of the original family and set up their own households[36]. Until recently, there used to be two different releases of PSID data, Release I (also known as Early Release) and Release II (also known as Final Release). Early release data were available for all years; final release data are available (at time of writing) only between 1968 and 1993. The variables needed for our study are available in both releases. The difference is that Release II data tend to be more polished and contain additional constructed variables. We use Release II data for the period 1968-1993 and Release I data for the period 1994-2001[37].

Because of successive improvements in Computer Assisted Telephone Interviewing (CATI) software,

[36] A distinction between original sample individuals, including their offspring if born into a responding panel family during the course of the study (i.e., both those born to or adopted by a sample individual), and nonsample individuals must be made. Details about the observations on non-sample persons and their associated weights and relevance are included in the appendix.

[37] We also have results obtained from a reduced sample using only Release I data for 1968-1993: estimates of the parameters of interest don't substantially differ from the full sample estimates.

the quality of the Public Release I files improved dramatically in recent waves, allowing the use of these data with confidence. The differentiation between Public Release I and Public Release II has recently been dropped altogether.

A.1 Sample selection

Unequal probabilities of selection were introduced at the beginning of the PSID (1968) when the original Office of Economic Opportunity (OEO) sample of poor families was combined with a new equal probability national sample of households selected from the Survey Research Center 1960 National Sample. Compensatory weights were developed in 1968 to account for the different sampling rates used to select the OEO and SRC components of the PSID.

The probability sample of individuals defined by the original 1968 sample of PSID families was then followed in subsequent years. A distinction between original sample individuals, including their offspring if born into a responding panel family during the course of the study (i.e., both those born to or adopted by a sample individual), and nonsample individuals was also made. Only original sample persons and their offspring have been followed. These individuals are referred to as sample persons and assigned person numbers in a unique range. If other individuals resided with the sample individuals, either in original family units or in newly created family units, data were collected about them as heads, spouses/long term cohabitors or other family unit members, in order to obtain a complete picture of the economic unit represented by the family. However, these nonsample individuals were not followed if they left a PSID family.

Step-by-step Sample Selection (data from 1967 to 2000). After dropping 10,607 individuals belonging to the Latino sample and 2263 individuals belonging to the new immigrant families added in 1997 and 1999, the joint 1967-2001 sample contains 50,625 individuals. After selecting only the observations on household heads we are left with 19,583 individuals.Dropping people younger than 25 or older than 65 leaves us with 18,186 people. Dropping the self employment observations leaves 14,866 persons in the sample. We then select only the individuals with at least 8 (possibly non continuous) observations, which further reduces the people in the sample to 6228. Dropping individuals with inconsistent education records leaves 6213 people in sample. Dropping individuals with missing, top-coded or zero earnings reduces the sample to 5671 individuals and dropping those with zero, missing or more than 5840 annual work hours brings the sample size to 5,660 individuals. We eliminate individuals with

year	Number of Observations	year	Number of Observations
1967	933	1983	1,775
1968	1,015	1984	1,802
1969	1,109	1985	1,808
1970	1,181	1986	1,829
1971	1,294	1987	1,837
1972	1,395	1988	1,840
1973	1,508	1989	1,838
1974	1,543	1990	1,809
1975	1,601	1991	1,780
1976	1,635	1992	1,697
1977	1,685	1993	1,698
1978	1,705	1994	1,638
1979	1,737	1995	1,588
1980	1,755	1996	1,510
1981	1,734	1998	1,425
1982	1,718	2000	1,298

Table 30: Number of individual observations, by year

Years of Education	Number of Individuals	Number of Observations
	Sample from 1967 to 2000	
less than 12	430	6,546
12 to 15	1,792	29,229
16 or more	863	14,945
	Sample from 1967 to 1996	
less than 12	430	6,380
12 to 15	1792	27,583
16 or more	863	14,034

Table 31: Distribution of individuals and total observations by years of education.

outlying earning records, defined as changes in log-earnings larger than 4 or less than -2, which leaves 5,477 individuals in the sample. Finally, dropping people connected with the SEO sample reduces the number of individuals to 3,085, with a total number of observations of 50,720.

The composition of the sample by year and by education group is reported in the following tables.

B CPS Data

the Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics.[38] This monthly survey of households is conducted for BLS by the Bureau of the Census through a scientifically selected sample designed to represent the civilian noninstitutional population. Respondents are interviewed to obtain information about the employment status of each member of the household 15 years of age and older. Each month about 50,000 occupied units are eligible for interview. Some 3,200 of these households are contacted

[38]The survey has been conducted for more than 50 years. Statistics on the employment status of the population and related data are compiled by the Bureau Labor Statistics (BLS) using data from the Current Population Survey (CPS).

Descriptive Statistics: Demographic Characteristics

	1967	*1980*	*1993*
age	38.04	38.93	44.29
average hourly wage-LTHS	11.82	12.69	11.02
average hourly wage-HSG	14.53	15.33	15.63
average hourly wage-CG	20.54	19.34	25.02
average ln(hourly wage)-LTHS	2.36	2.42	2.24
average ln(hourly wage)-HSG	2.59	2.59	2.60
average ln(hourly wage)-CG	2.90	2.84	3.04
median hourly wage-LTHS	10.59	11.61	10.31
median hourly wage-HSG	14.09	13.96	13.61
median hourly wage-CG	19.24	17.45	21.46
average hours worked-LTHS	2246	2093	2121
average hours worked-HSG	2268	2125	2155
average hours worked-CG	2234	2158	2196

Table 32: PSID Sample Descriptive Statistics: Earnings are annual earnings and hours are annual hours worked. Wages are hourly wages computed as annual earnings divided by annual hours worked. Both wages and earnings are expressed in 1992 dollars.

but interviews are not obtained because the occupants are not at home after repeated calls or are unavailable for other reasons. This represents a non-interview rate for the survey that ranges between 6 and 7 percent. In addition to the 50,000 occupied units, there are 9,000 sample units in an average month which are visited but found to be vacant or otherwise not eligible for enumeration. Part of the sample is changed each month. The rotation plan, as explained later, provides for three-fourths of the sample to be common from one month to the next, and one-half to be common with the same month a year earlier. The CPS has been used to collect annual income data since 1948, when only two supplementary questions were asked in April: "How much did ... earn in wages and salaries in 1947 ..." and "how much income from all sources did ... receive in 1947". Over the years, the number of income questions has expanded, questions on work experience and other characteristics have been added, and the month of interview relating to previous year income and earnings has moved to March. This yearly survey goes under the name of March CPS Supplement.[39] Age classification is based on the age of the person at his/her last birthday. The adult universe (i.e., population of marriageable age) is comprised of persons 15 years old and over for March supplement data and for CPS labor force data. Each household and person has a weight that should be used in producing population-level statistics. The weight reflects the probability sampling process and estimation procedures designed to account for nonresponse and undercoverage. Unweighted counts can be very misleading and should not be used in demographic or labor force analysis.

[39]Today, information is gathered on more than 50 different sources of income, including noncash income sources such as food stamps, school lunch program, employer-provided pension plan and personal health insurance. Comprehensive work experience information is given on the employment status, occupation, and industry of persons 15 years old and over.

Shares (%) of workers by years of schooling in each sample year

	Years of schooling				*Years of schooling*		
year	**<12**	**12-15**	**>15**	*year*	**<12**	**12-15**	**>15**
1968	41.4	47.5	11.1	1985	20.0	60.0	20.0
1969	40.0	48.7	11.3	1986	19.0	60.0	21.0
1970	38.4	50.1	11.5	1987	19.0	60.0	21.0
1971	36.7	51.1	12.2	1988	18.7	60.0	21.4
1972	35.3	51.9	12.8	1989	18.2	59.8	22.0
1973	33.6	53.1	13.3	1990	17.5	60.1	22.4
1974	32.8	53.4	13.8	1991	16.8	60.6	22.6
1975	31.3	54.1	14.6	1992	15.7	61.7	22.6
1976	29.2	55.2	15.6	1993	14.8	61.7	23.4
1977	28.8	55.2	16.0	1994	14.7	61.8	23.5
1978	28.1	55.8	16.1	1995	14.6	61.1	24.3
1979	26.3	57.0	16.7	1996	14.8	60.5	24.7
1980	25.1	57.6	17.3	1997	14.7	60.7	24.6
1981	23.8	58.6	17.6	1998	14.6	60.2	25.2
1982	22.8	58.8	18.4	1999	14.4	59.8	25.8
1983	21.0	59.3	19.7	2000	14.1	59.8	26.1
1984	20.3	59.6	20.1	2001	13.8	59.6	26.6

Table 33: Percentage of people without High School degree (less than 12 years of schooling), with High School degree (12-15 years of schooling) and with College degree (at least 16 years of completed schooling), for the years 1968-2001. Based on CPS March Supplement data. The sample consists of workers only.

Sample selection. We use the March CPS yearly files and additional files from 1968 to 2001. We use the CPI for all urban consumer (with base year 1992) to deflate the CPS earning data and drop all observations that have missing or zero earnings. Since the earning data are top-coded for confidentiality issues until 1995, we have extrapolated the average of the top-coded values by using a tail approximations based on a Pareto distribution.[40] For the period 1996-2000 BLS provides the averages of unreported values for different age/sex/empl. groups. Therefore the overall mean of the distribution is easy to recover and we do not use any tail adjustment.

The averages of earnings, both censored and tail-adjusted, are reported in table (B), categorized by education group and year.

[40]This procedure is based on a general approach to inference about the tail of a distribution originally developed by Hill (1975). This approach has been proposed as an effective way to approximate the mean of top-coded CPS earning data by West (1985); Polivka (2000) provides evidence that this method closely approximates the average of the top-coded tails by validating the fitted data through undisclosed and confidential non top-coded data available only at the BLS.

year	Average Earnings (yearly, in $)						year	Average Earnings (yearly, in $)					
	censored data			adjusted data				censored data			adjusted data		
	<12	12-15	>15	<12	12-15	>15		<12	12-15	>15	<12	12-15	>15
1968	14336	20609	35324	14384	20688	35725	1985	11676	19319	34656	11762	19463	36276
1969	14794	21180	35432	14794	21261	35780	1986	11626	19682	35301	11779	19834	37290
1970	14903	21610	37036	14924	21702	37540	1987	11820	20139	36598	12002	20322	39057
1971	14690	21417	36364	14839	21556	36863	1988	11991	20364	36049	12229	20668	38349
1972	14841	21157	36340	14920	21242	37115	1989	11755	20288	36084	12152	20689	38472
1973	15383	22249	37538	15541	22469	38507	1990	11265	20334	36405	11369	20811	39571
1974	15020	22232	37057	15118	22435	38110	1991	11127	19733	35296	11230	20084	38467
1975	14208	21199	34357	14335	21405	35595	1992	10661	19381	35210	10777	19710	38507
1976	13333	20468	34367	13400	20689	35949	1993	10446	19422	35363	10597	19784	39089
1977	13675	20584	34544	13796	20863	36467	1994	9955	19167	35610	10117	19617	40228
1978	13396	20979	34810	13484	21272	37267	1995	10227	19439	35991	10468	19913	41664
1979	13395	21121	34499	13613	21409	37503	1996	10579	20306	39643	10579	20306	39643
1980	13513	20678	33632	13692	21040	37283	1997	10790	20586	40073	10790	20586	40073
1981	12600	19625	32169	12701	20077	36040	1998	11287	20809	41019	11287	20809	41019
1982	12213	19223	32311	12337	19410	34289	1999	10984	21406	42794	10984	21406	42794
1983	11699	18746	32621	11730	19000	35209	2000	10784	21626	42148	10784	21626	42148
1984	11796	18928	33191	11906	19178	35585	2001	11330	22346	45147	11330	22346	45147

Table 34: Average earnings by education group, all years. Earnings are in 1992 dollars. Censored averages are based on unadjusted, top-coded data. Adjusted averages are based on data with Pareto-tail adjustment. From 1996 onwards censored and adjusted averages correspond because of larger data disclosure.

Calibrated Parameter Values for Benchmark

Parameter	*Value*	*Moment to Match*
$\bar{j}$	79	Maximum lifespan after labor market entry
j_r	50	Maximum years of working life
$\{\lambda_j\}$	-	Survival rates
β	0.967	Wealth-Income ratio excluding top 1%
$\underline{l}$	0.74	Average duration of unemployment
q	.175	Unemployment incidence
$sub_{high\ school}$	.019	Direct tuition cost of High School
$sub_{university}$	.19	Direct tuition cost of College
$\overline{a}$	-.865	Fraction of households with net worth ≤ 0
α	.35	Capital share in total output
δ	.065	Depreciation rate
p	.307	Pension replacement rate
t_l	.27	Labor income tax
t_K	.40	Capital income tax
χ	-.32	Aggregate property crime rate
$\overline{u}$	-.80	Elasticity of crime rate w.r.t. expected prison term

Table 35: Value of Parameters Calibrated in Benchmark

Quasi-linear utility terms associated to education

	Education group	
Ability Bin	*High School*	*College*
Bin 1	0.27	-5.80
Bin 2	0.39	-4.32
Bin 3	0.49	-2.55
Bin 4	0.89	-0.64

Table 36: Quasi-linear utility terms associated to being in education for High-School and College. The terms differ by ability bin because the enrolment rates that are matched are different in each bin.