Essays on Equilibrium Policy Analysis

Giovanni Gallipoli

A Dissertation submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
University College London

London
September 2006

Abstract

This thesis describes and implements a method to carry out policy analysis within an equilibrium framework. This method allows to account for potential effects induced by price adjustments. The analysis is based on overlapping generation, life-cycle models where heterogeneous agents make endogenous decisions regarding their consumption and education as well as labour supply and criminal activity. Some of the agent's optimising decisions (education, crime) are discrete choices. The first part of the thesis deals with the issue of including binary decision choices in a life cycle model: the implications of non-convex choice sets in life cycle models with uninsurable idiosyncratic risk are studied in detail and some results on the properties of the individual problem's solution are provided. Next, we apply the proposed framework to analyse two distinct policy questions. The first application looks at the equilibrium analysis of tuition policies on the distribution of education and income. Empirical evidence suggests a link between human capital accumulation and wage dispersion. We experiment with college tuition subsidies and find that while in partial equilibrium such policies can be very effective in increasing education levels and reducing inequality in general equilibrium the results are less encouraging. The second application considers 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. I find that policies targeting crime reduction through increases in high school graduation rates are cost-effective, especially if they are targeted at the poor.

Contents

List of Figures

List of Tables

Acknowledgements

I have a debt of gratitude to my supervisors, Costas Meghir and Gianluca Violante, for their guidance, patience and support, both professional and personal. I am also particularly grateful to Lars Nesheim and Giulio Fella, who proved to be dear friends as well as valuable colleagues.

Over the past few years I have greatly benefitted from conversations with Orazio Attansio, Monica Costa-Dias, Hamish Low, Nicola Pavoni and Victor Rios-Rull. I have learnt a good deal from those conversations. I also owe much to the researchers at IFS: it was a pleasure to work there, in an environment which was both friendly and stimulating.

Among the many friends I have met during the course of my PhD, a special mention should go to Robin, Mario, Ben, Elena, Renata and Matt, as well as Makoto, Josep and Marco (a group of young economists, also known as the "suinettos"). I consider myself lucky to have met them.

Dragana has been an invaluable source of support over the years and I owe her much more than words can say. My whole family, even in moments of extreme distress, has never failed me.

This thesis is dedicated to my father, a honest and simple man, who would have enjoyed immensely seeing it in print.

Declaration

1. No part of this thesis has been presented to any University for any degree.
2. Chapter 3 was undertaken as joint work with Costas Meghir and Gianluca Violante.
3. Chapter 4 was undertaken as joint work with Giulio Fella.

Giovanni Gallipoli

Chapter 1

Introduction

The chapters of this thesis analyse elements of individual life-cycle decisions including consumption, work, education, saving and criminal activity within a structural equilibrium framework. Rational agents make optimal decisions based on available resources and objectives, as in Modigliani and Brumberg (1954) and Friedman (1957). They also react to changes in the economic environment and policy regime. The aggregation of individual choices generates distributions of economic outcomes which are computed and studied in detail.

As pointed out by Cunha, Heckman, and Navarro (2005), most studies focussing on the distribution of economic outcomes such as earnings, consumption and investment are descriptive in nature. In order to understand distributional effects and do policy analysis it is useful to move beyond descriptive statistics and build counterfactual outcomes. The standard platform of analysis in the policy evaluation field has been represented by randomised experiments which naturally or artificially provide a comparison between a treatment and a control group (for a discussion see Heckman (1992)). Meghir (2006) points out that randomisation and quasi-experimental methods can offer only limited answers to the questions of policy design and there is ample scope to complement them with alternative methods. The limitations of standard randomisation methods stem from different aspects of the evaluation problem. For example, it may be the case that the implementation of a programme induces changes in individual behaviour which bear effects in the long-term and are hardly captured by short-term comparisons of treatment and control group (see Heckman, Lochner, and Cossa (2002), and Jerome, Dustmann, Meghir, and Robin (2006), for examples and discussions of this issue). Methods that are able to approximate long-term outcomes are a valuable addition to short-term analysis. A second, serious problem is that a policy intervention can have effects on individuals who are not directly targeted. This is a violation of the basic identification assumption which allows to distinguish between treatment and control groups. Angelucci and DeGiorgi (2006) find evidence of such violation using experimental data from Mexico. Finally, while small scale interventions can have negligible effects on prices, the same is not necessarily true of broad programmes characterized by large take-up rates, (see Heckman, Lochner, and Taber (1998a, 1998b, 1998c), Alonso-Borrego, Fernandez-Villaverde, and Galdon-Sanchez (2004), Lee and Wolpin (2006)). If policy interventions simultaneously alter the behaviour of many agents they can induce changes in the economic environment, for example by determining new levels of aggregate labor supply or capital accumulation. In this way, policies targeting the wider population potentially induce variation in market prices which can account for a substantial portion of the final outcome.

The programme evaluation literature often assumes a simple dichotomy between treatment and control: agents are usually split into two groups, treated and non treated. Being part of a treatment group is often associated to an agent's optimal take-up decision as well as some assignment rule, and it can depend on discrete choices like being part of the workforce or studying. Modelling binary choices presents analytical and computational challenges because of non-convexities in the choice sets.

The first part of this thesis concerns the effect of introducing non-convex choice sets in a standard life-cycle maximisation problem, see Townsend and Ueda (2003) and Gomes, Greenwood, and Rebelo (2001). The optimisation problem faced by the agents is spelled in detail within a life cycle model with uninsurable idiosyncratic risk. We discuss the implications of non-convex choice sets and some general properties of non concave value functions, with a focus on issues of non-uniqueness and continuity of individual optimal policies. We present conditions under which the individual life-cycle problem can be solved in sequential form and prove existence of an optimal plan.

Proof is provided that if a functional equation exists, then it achieves the optimum value of the sequence plan. Finally, some properties of the optimal binary choice at relevant state-space locations are described.

After discussing the peculiarities of models with binary decisions the thesis shifts its focus on the implementation of counterfactual analysis for policy evaluation. Two different General Equilibrium applications are presented. In the first one a model of education, life-cycle labour supply and consumption under uncertainty is specified to investigate both short and long term consequences of policies designed to enhance investment in Human Capital. The impact of a tuition subsidy, both for the aggregate economy and for agents of different ability and wealth, is analysed.

The second application compares the effectiveness of alternative policies targeting a reduction in property crime rates. A model with endogenous consumption, saving, education and crime decisions is used and different counterfactual scenarios are constructed in order to identify the effectiveness and cost of different interventions. The reaction to the policy of different types of agents is also studied.

Methods which allow to build counterfactual outcomes (that is, outcomes not directly observed in data) provide a way to move beyond aggregate measurements of policy effects and to establish whether and to what extent different people in an initial distribution are affected by some policy. Cunha, Heckman, and Navarro(2004, 2005) stress that, unless available panel data are good enough to build alternative outcomes in the presence of policy reforms, only counterfactual methods can provide answers on how different people "benefit or lose, how much they lose, how they would vote in advance of the reform and how they would vote after it is implemented, once the ex ante uncertainty surrounding the outcomes of the reform is resolved".

In both applications the aggregate and distributional effects of the different programmes are presented. It becomes apparent that aggregate measures provide only a rough approximation of the effects of a programme for most individuals. The policy effects are extremely strong for some groups and almost non-existent for others. Moreover, the short-term, off-equilibrium effects are very different from the long-term, equilibrium effects.

Chapter 2

On Non-Convexities in a Life-Cycle Model

2.1 Introduction

Studies on topics such as education, retirement, unemployment and portfolio decisions often include binary decision variables on which budget sets depend. For example, Gomes, Greenwood, and Rebelo (2001) study an infinite horizon search-theoretic model of equilibrium unemployment in which the employment decision is a binary variable. They find that the outer envelope of the value of being employed or searching is not concave due to the depression where these two conditional values intersect over the assets' domain. This kind of "butterfly" value functions are a common event in the presence of binary choices. This chapter discusses the issue of non-convexities implied by the presence of binary choice variables in a dynamic programming context. Concavity of value functions is a highly desirable thing as it guarantees differentiability, sufficiency of the first order conditions for a maximum and applicability of the Envelope Theorem. Gomes, Greenwood, and Rebelo (2001) concavize the value function by including a normally distributed shock that "fills up" the convex hull generated by the "butterfly" crossing. This method is often effective in models with continuous future shocks but no general result guarantees the effectiveness of this method under all circumstances[1].

We focus on a finite life cycle model. We prove our results for a model with binary, repeated education choices. After solving the problem in sequential form and proving the existence of an optimal plan achieving a sequential maximum, we show that the maximum can be obtained by a functional equation and that such functional equation does exist. We study the functional equation in both its conditional and unconditional forms and show that the individual optimal policy (the pair of education and asset saving decisions) is piece-wise continuous and, even without an explicit tie-break assumption, can be considered for practical purposes as being single-valued. Using an argument proposed by Pavoni (2006), we notice that the points where the discontinuities occur are the same points at which the policies are not single-valued, but such points are never chosen.

2.2 Literature

Within the infinite horizon models' family, Townsend and Ueda (2003) study a dynamic model with financial deepening, intended in the sense of a binary portfolio choice carrying a one-off fixed cost of entry. They present proofs of existence of an optimal program and its equivalence with the value function approach. [2] Our work is in the spirit of that of Townsend and Ueda (2003), and whenever possible uses notation and procedures similar to theirs. However, some of the properties of the finite lifetime problem are different and new results are provided for this case.

In the analysis of optimal unemployment benefit contracts Pavoni (2006) discusses the case of a non concave value function emerging from the envelope of two conditional value functions. He studies the 'switching points' (utility levels at which the upper

[1] A simple argument can show that not all continuous shocks are able to make the unconditional value concave: consider a shock process which can be fully described by its finite mean and variance and which is able to concavize the envelope of two conditional value functions. If we let the variance of such shock become progressively smaller, there might be a positive variance below which the shock process will not vary enough to smooth the unconditional envelope. In other words, shifts in the conditional values induced by small enough shocks can be unable to generate a smoothing of the unconditional envelope.

[2] Townsend and Ueda (2003) also show that an optimal portfolio choice can convexify a non-convex participation technology.

envelope function switches between two different conditional functions) and finds that, for the class of concave and continuously differentiable functions, each switching point possesses a very nice characteristic: the 'kink' at the switching point is an 'inward' one. It can be shown that this implies that the optimal choice of the continuation utility can never be at a switching point. We will also make use of this argument.

2.3 A model of education choices

In this chapter we derive the optimal consumption and schooling choices for an individual of given ability who supplies labor in a competitive market. A unique good is produced in the economy, and it can be either consumed or saved as an asset to carry over to the future. Different kinds of education command different returns. Wage differences among persons are generated by differences in education (between group inequality) and by differences in ability and labor efficiency (within group inequality).

We use the index $j \in \Upsilon \equiv \{1, 2, .., .., \overline{j}\}$ to indicate agents'age. Agents have an age-related probability to die measured by $(1 - s_j)$, with $(1 - s_j) < (1 - s_{j+1})$, $(1 - s_0) = 0$ and $\left(1 - s_{\overline{j}}\right) = 1$.

2.3.1 Demographics and preferences

Each agent's life starts at age 1 and lasts at most $\overline{j}$ periods, after which death is certain.[3] Agents are faced with educational choices at the beginning of their lives and base such choices on returns and costs of education and on their initial asset holding, ability and idiosyncratic shocks. Over the life cycle they choose the labor supply path that maximizes their expected lifetime utility.

Let e denote the individual educational attainment, $e \in \Im \equiv \{e_1, e_2, e_3\}$, with $e = e_1$ the lowest and $e = e_3$ the highest. Also, denote individual ability by $\theta \in [\theta_{\min}, \theta_{\max}] \equiv \Theta$ and let $\{z\}_{j=1}^{\overline{j}}$ be a sequence of uninsurable idiosyncratic shocks, $z_j \in Z$; finally, use $a_j \in \bar{A}$ to denote individual asset holdings at age j.

Assumption 1 *Θ, $\Im$, Z and $\bar{A}$ are compact sets. $\bar{A}$ is a convex set.*

[3] Accidental bequests can be left at the end of life.

Define the log of individual labor efficiency as

$$\epsilon_j(\theta, e, z) = \theta + \xi_j(e) + z_j \tag{2.1}$$

where $\xi_j(e)$ is an age-dependent trend in efficiency specific to each education group $e \in \Im$. The market wage for education level e is denoted as w_e and individual labor supply as $n \in [0,1]$. Leisure, $l = (1-n)$, is a choice variable when employed; however we assume that the amount of leisure is not a choice variable for students but an increasing function of ability, which is defined as $f^S = f^S(\theta) \in [0,1]$.

Agents pay proportional taxes τ_{n^e} and τ_k on, respectively, labor and asset income. Labor factor prices w_e, $e \in \Im$, and capital factor price r are exogenous and constant.

Individual consumption at age j is denoted as $c_j \in \Re_+$; direct cost of schooling as D_e, government subsidies towards education as Tr_e, intertemporal discount factor as $\beta > 0$ and individual bequests received by an household at age j as q_j.

The period utility $u(c,l)$ varies with consumption c and leisure $l = (1-n)$, and is defined as $u : \Re_+ \times [0,1] \longrightarrow \Re$. We make the following assumptions on $u(c,l)$:

Assumption 2 *$u(\cdot)$ is (at least) twice continuously differentiable on its domain, strictly increasing and strictly concave in both its arguments and satisfies the Inada conditions, $\lim_{c \longrightarrow 0} u_c = \infty$, $\lim_{c \longrightarrow \infty} u_c = 0$ and $\lim_{l \longrightarrow 0} u_l = \infty$, so that consumption is never zero, savings never equal available resources, and agents always consume a minimum amount of leisure.*

It follows that the domain and range of $u(\cdot)$ are $\Re_{++} \times (0,1]$ and $\Re$, respectively.

2.3.2 Constraints

Several constraints restrict the behaviour of agents.

Assumption 3 *$a_j \geq a_{\min}$ for every j and $a_{\bar{j}+1} \geq 0$.*

The first inequality is a standard borrowing constraint imposing a lower bound on asset holding. In what follows we consider the case of $a_{\min} = 0$.

The second inequality is a terminal condition for agents reaching age $\bar{j}$: they can consume all their assets before kicking the bucket but are barred from dying in debt.

We also assume a law of motion for the transitory idiosyncratic shock z.

Assumption 4 *Let Z be a compact (Borel) set in $\Re$, and let π be a transition function on the measurable space $(Z, \digamma(Z))$, denoted as $\pi_{z_{j+1}|z_j} = \pi\{z_{j+1} \mid z_j\}$, with $z_1 = \bar{z}_1$. The function $\pi(\cdot)$ has the Feller property and is monotone*[4].

Sequential problem: given some initial conditions $\bar{x}_1$, an age 1 agent's utility over sequences of consumption and leisure, $c = \left\{c_1, ..., c_{\bar{j}}\right\}$ and $l = \left\{l_1, ..., l_{\bar{j}}\right\}$, is denoted as $U(\bar{x}_1, c, l)$ and can be written as the expected discounted sum of period utilities

$$U(\bar{x}_1, c, l) = E_{z \in Z} \sum_{j=1}^{\bar{j}} \left(\prod_{i=1}^{j} s_i\right) \beta^{j-1} u\left(c_j(z_j), l_j(z_j)\right) \tag{2.2}$$

Such utility is allowed to take values over the set identified by the following restrictions

$$c_j + a_{j+1} = \left[1 + r(1 - \tau_k)\right] a_j + w_e \exp^{\epsilon_j} n_j (1 - \tau_{n^e})(1 - \mathrm{d}_j) - (D_e - Tr_e)\, \mathrm{d}_j \tag{2.3}$$

where d_j is a binary variable which is 1 if the agent is in education and 0 otherwise.

2.4 The agent's problem

At the beginning of life individuals decide whether to engage in education or go to work.

Assumption 5 *Agents who terminate education and start working cannot go back to education at a later age.*

The above assumption can be removed allowing for agents to go back to education: of course things become more complicated as agents which experience bad labor shocks

[4]The Feller property holds if and only if the expectation operator using the transition function $\pi_{z_{j+1}|z_j} = \pi\{z_{j+1} \mid z_j\}$ on $(Z, \digamma(Z))$ maps the space of continuous bounded functions on Z into itself. Monotonicity means that for every nondecreasing function $f : Z \longrightarrow R$ the expectation of f is also nondecreasing.

are more likely to go back to school. However, it could be proven that there is no incentive to go back to education beyond a certain age.

The education choice made by each agent depends on the relative returns to different education levels in the economy, the pecuniary costs of schooling D_e and government education subsidy Tr_e, the innate type θ, the current idiosyncratic shock z, and the current asset holdings a.

Assumption 6 *(Education Progress) To pass from education level e_1 to education level e_2 an agent has to stay in school from age 1 to age $j(2)$. Similarly, to pass from e_2 to e_3 an agent has to stay in school from age $j(2)+1$ to age $j(3)$. No schooling is possible after age $j(3)$.*

Involuntary unemployment is not possible in this model.

Assumption 7 *All agents who decide to work can find employment at the existing market wages.*

Given some skill prices w_e, direct costs of schooling D_e and government education subsidy Tr_e, the binary function $\mathrm{d}_j : \Upsilon \times \Theta \times \Im \times Z \times \bar{A} \longrightarrow \{0,1\}_j$, $j \in \{1, j(3)\}$ describes an agent's educational choice as a mapping from the space of age, ability, education, idiosyncratic shocks and assets into the age j employment set $\{0,1\}_j$, where 1 stands for full-time education status and 0 stands for full-time worker[5].

We define educational attainment as a function $e : \cap_{j=1}^{j(3)} \{0,1\} \longrightarrow \Im$, and we denote it as

$$
\begin{aligned}
e &= e_1 \quad if \quad \sum_{j=1}^{j(3)} \mathrm{d}_j < j(2) \\
e &= e_2 \quad if \;\; j(2) \leq \sum_{j=1}^{j(3)} \mathrm{d}_j < j(3) \\
e &= e_3 \quad if \;\; j(3) = \sum_{j=1}^{j(3)} \mathrm{d}_j
\end{aligned}
$$

[5]Denote $x = (\theta, e, z, a) \in \digamma(X)$, where $\digamma(X) = \digamma(\Theta) \times \digamma(\Im) \times \digamma(Z) \times \digamma(\bar{A})$ is the σ algebra on $X = (\Theta \times \Im \times Z \times \bar{A})$, then $\mathrm{d}_j = \mathrm{d}_j(x)$.

Such educational sorting mechanism hinges on the assumption that progress from one educational level to the next may require more than just one study spell.

Remark 1 *Incomplete study spells do not change the educational status* (e_i) *of an agent.*

Remark 2 *Conditional on the choice of entering the labor market, the labor supply policy of an agent is* $n_j = n_j(\theta, e, z, a \mid \mathrm{d}_j = 0)$. *Using the intra-temporal margin condition it is possible to express the individual labor supply as a function of optimal consumption and real wage*[6].

Without conditioning on the current education decision, the optimal policy of an agent can be represented as a vector $p_j = (\mathrm{d}_j, a_{j+1})$, where a_{j+1} is the optimal saving policy and d_j the binary education decision. The policy space P is

$$p_j = (\mathrm{d}_j, a_{j+1}) \in P \equiv \{0,1\} \times \Re_+$$

Life-long profiles for asset holding are uniquely determined by ability, education, past asset holdings and individual shocks. Given $x_j = (\theta, e_j, z_j, a_j)$ and d_j, we can write $a_{j+1} = a_{j+1}(x_j)$ for $j = 1, ..., \bar{j}$. We assume that initial asset levels are finite and positive, $0 < a_1 < \infty$, and since a_{j+1} cannot exceed available resources we have that for all j[7]

$$a_{j+1} \leq Ra_j + \tilde{w}n_j(1 - \mathrm{d}_j) - (D_e - Tr_e)\,\mathrm{d}_j \tag{2.4}$$

The age j policies that are feasible when $\mathrm{d}_{j-1} = 1$ are

$$p_j \in \Gamma(x_j, p_{j-1}) \equiv \{0,1\} \times [0, Ra_j + \tilde{w}n_j(1 - \mathrm{d}_j) - (D_e - Tr_e)\,\mathrm{d}_j] \tag{2.5}$$

The set of feasible policies Γ is nonempty and compact valued. When $\mathrm{d}_{j-1} = 1$, Γ is a correspondence defined as $\Gamma : \Theta \times \Im \times Z \times \bar{A} \longrightarrow \{0,1\} \times \Re_+$. In this case Γ is generally non convex valued because of the binary choice $\mathrm{d}_j(x_j)$. However, Γ

[6] The analytical details of the labour/leisure inter-temporal choice are provided in the Appendix.

[7] Here $R = [1 + r(1 - \tau_k)]$ and $\tilde{w} = w_e \exp^{\epsilon_j}(1 - \tau_{n^s})$. Incidentally, we assume $\tilde{w}$ is finite and positive.

is convex conditional on the implicitly determined $\mathrm{d}_j(x_j)$, that is $\Gamma(x_j, \mathrm{d}_{j-1} = 1) = 1 \times [0, Ra_j - (D_e - Tr_e)]$ if $\mathrm{d}_j(x_j) = 1$ and $\Gamma(x_j, \mathrm{d}_{j-1} = 1) = 0 \times [0, Ra_j + \hat{w}n_j]$ if $\mathrm{d}_j(x_j) = 0$, which are both convex sets.

When $\mathrm{d}_{j-1} = 0$, Γ is convex-valued: in fact, by assumption 5, the set of feasible policies can be reduced to:

$$p_j \in \Gamma(x_j, \mathrm{d}_{j-1} = 0) \equiv 0 \times [0, Ra_j + \hat{w}n_j]$$

and Γ is a convex-valued correspondence defined as $\Gamma : \Theta \times \Im \times Z \times \bar{A} \longrightarrow 0 \times \Re_+$.

The resource constraint implies that, given a value for $p_j = (\mathrm{d}_j, a_{j+1})$ and $p_{j-1} = (\mathrm{d}_{j-1}, a_j)$, consumption at j can be written as[8]

$$c_j(p_j, p_{j-1}) = Ra_j + \widetilde{w}n_j(1 - \mathrm{d}_j) - (D_e - Tr_e)\,\mathrm{d}_j - a_{j+1} \tag{2.6}$$

Using (2.6) we obtain the indirect instantaneous utility

$$\mathrm{v}(p_j, x_j, p_{j-1}) \equiv u(c_j(\mathrm{d}_j, \mathrm{d}_{j-1}, a_{j+1}, x_j)) \tag{2.7}$$

Assumption 8 $\mathrm{v} : P \times \Delta \times P \longrightarrow \Re_+$, *as defined in (2.7), is a measurable function*[9]*.*

The indirect utility function $\breve{U}$, given initial conditions $\bar{x}_1$ and policy sequence $p = \left\{p_1, ..., p_{\bar{j}}\right\}$, is defined as the age 1 discounted sum of indirect instantaneous utilities

$$\breve{U}(\bar{x}_1, p) = E_{z \in Z} \sum_{j=1}^{\bar{j}} \left(\prod_{i=1}^{j} s_i \right) \beta^{j-1} \mathrm{v}(p_j, x_j, p_{j-1}) \tag{2.8}$$

where $\breve{U} : P^{\bar{j}} \longrightarrow \Re$, where $P^{\bar{j}}$ is the $\bar{j}-$*product* of the set of policies P.

The household problem can then be stated as

$$U^*(\bar{x}_1) = \sup_{p \in B(\bar{x}_1)} \breve{U}(\bar{x}_1, p) \tag{2.9}$$

[8] We are using the short notation $\tilde{w} = w_e(1 - \tau_{n^e})\exp^{\epsilon_j}$.

[9] Measurable on the $\mathcal{F}(P) \times \mathcal{F}(\Delta) \times \mathcal{F}(P)$ σ-algebra. Recall that $x_j \in \Delta = \Upsilon \times \Theta \times \Im \times Z \times \bar{A}$.

where $B(\bar{x}_1)$ identifies the set of feasible policy sequences

$$B(\bar{x}_1) \equiv \{p_j \in \Gamma(x_j, p_{j-1})\} \tag{2.10}$$

2.4.1 Optimality

In this section we claim that an optimal plan exists and discuss the value function representation of this problem . The relevant proofs are presented in the Appendix.

Lemma 1 *The set $B(\bar{x}_1)$ is compact in the $\overline{j}$ − product topology.*

If the supremum in (2.9) is feasible, then it is also a maximum and the associated policy plan p^* is the optimal plan.

However, before stating that the supremum operator can be replaced with a maximum, we restrict the individual's utility $U(\bar{x}_1, c, l)$ defined in equation (3.1) to be bounded from above.

Proposition 1 $U(\bar{x}_1, c, l) < \infty$.

Now we can state the existence of an optimal plan.

Proposition 2 *There exists an optimal plan p^* such that $\breve{U}(\bar{x}_1, p^*)$ as defined in equation (2.8) is equal to the supremum $U^*(\bar{x}_1)$ as defined in equation (2.9).*

We use value functions to characterise the optimal path[10]. Next we show that a functional equation is an equivalent and unique approach to the household's sequence problem (2.9).

Lemma 2 *$U^*(\bar{x}_1)$ is continuous in $\bar{x}_1$, measurable in $\bar{x}_1$ and monotonically increasing with changes of θ, $\bar{z}_1$ and $\bar{a}_1$.*

[10] In this section we use an hyphen "′" to identify next period unknown values and often omit the age/time subscripts for notational simplicity.

The functional equation is then

$$J\left(x_j, p_{j-1}\right) = \sup_{p_j \in \Gamma(x_j, p_{j-1})} \mathrm{v}\left(p_j, x_j, p_{j-1}\right) + \left(s_j \beta^{j-1}\right) \int_Z \pi_{z_{j+1}|z_j} J\left(x_{j+1}, p_j\right) dz_{j+1} \tag{2.11}$$

for given initial condition $\bar{x}_1$.

As, from Lemma 2, $U^*(\bar{x}_1)$ is continuous and measurable in $\bar{x}_1$, J can be chosen from the functional space of continuous and measurable functions, denoted as $\wp$.

If there exists a function $J \in \wp$ that satisfies (2.11) then $J(\bar{x}_1)$ coincides with $U^*(\bar{x}_1)$ in (2.9). We first note that the demographics of this model imply that $J(x_j, p_j) = 0$ for all $j > \bar{j}$. Then, following Stokey, Lucas, and Prescott (1989), page 247, and Townsend and Ueda (2001), with slight modifications, we can prove that

Proposition 3 *The value of the functional $J \in \wp$ defined in (2.11) achieves the value of the sequence problem $U^*(\bar{x}_1)$ defined in (2.9).*

Corollary 1 *Given $\bar{x}_1$, the value of $J(\bar{x}_1)$ is unique and equivalent to $U^*(\bar{x}_1)$.*[11]

We also want to prove that not only the functional equation and the sequence problem return the same solution value, but that the associated optimal policies coincide.

Proposition 4 *The optimal policies associated to the function equation $J(\bar{x}_1)$ coincide with the optimal policies of the sequence problem $U^*(\bar{x}_1)$.*

Having shown the equivalence of the sequence solution and of the functional equation, we are left to show that a value function exists.

Proposition 5 *A value function $J^*(x_j, p_{j-1})$ satisfying the functional equation (2.11)*

[11] The value of $U^*(\bar{x}_1)$ is the supremum of $\breve{U}(\bar{x}_1, p)$ in (2.8) over the set of policies P, and the supremum of any function is unique. By proposition (3) $J(x_1)$ achieves $U^*(\bar{x}_1)$, and therefore $J(x_1)$ is also unique and equivalent to $U^*(\bar{x}_1)$.

exists. Moreover, the correspondence $G : \Delta \longrightarrow \Re_+$, defined by

$$G(x_j, p_{j-1}) = \{ p_j \in \Gamma(x_j, p_{j-1}) : J^*(x_j, p_{j-1}) = \\ = \mathrm{v}(p_j, x_j) + s_j \beta^{j-1} \int_Z \pi_{z_{j+1}|z_j} J^*(x_{j+1}, p_j)\, dz_{j+1} \} \tag{2.12}$$

is non-empty, compact-valued and upper hemi-continuous.

Proposition 5 states that the correspondence of optimal policy functions $G(x_j, p_{j-1})$ is non-empty, compact-valued and upper hemi-continuous. This result turns out to be extremely helpful.

2.4.2 Value functions

In this section we study the shape and properties of the value function $J^*(x_j, p_{j-1})$. We call $J^*(x_j, p_{j-1})$ the unconditional value function because it is defined over all the possible (current) education choices.

In order to fully characterize the unconditional value function it is helpful to study the two conditional value functions which are obtained by assigning a value to the (current) binary choice d_j; the conditional versions of $J^*(x_j, p_{j-1})$ are the value of employment when $\mathrm{d}_j = 0$, and the value of education when $\mathrm{d}_j = 1$.

We denote the *conditional* value function as $\bar{J}(x_j, p_{j-1} \mid \textit{condition})$, with the condition being the value of d_j[12].

The *unconditional functional* equation $J^*(x_j, p_{j-1})$ is the *upper envelope* of the conditional values of employment and education. We discuss the properties of such envelope and characterize the optimal policies.

Without loss of generality, we reduce the complexity of the value functions associated to different kinds of employment by making the choice of employment-type irreversible (this is equivalent to assume that the costs of reverting to different, feasible 'careers' are sufficiently high).

Assumption 9 *Working agents can find a job in the spot-market corresponding to their own education level or a lower one; this initial choice is irreversible.*

[12] Such notation allows to summarize education status for the last 2 periods (d_{j-1} and d_j).

We start our discussion by examining the conditional value of employment, that we call W_j.

Lemma 3 *Given assumption 5, the conditional value of employment $W_j(\theta, e, z, a)$ exists, is unique and is defined as*

$$\begin{aligned} \bar{J}\left(x_j, p_{j-1} \mid \mathrm{d}_j = 0\right) &= W_j(\theta, e, z, a) = \\ &= \max_{a', n} \; u\left(c, 1-n\right) + s_j \beta \int_Z \pi_{z'|z} \, W_{j+1}\left(\theta, e, z', a'\right) \, dz' \end{aligned} \tag{2.13}$$

Remark 3 *By Assumptions 6 and 9, $W_j\left(\cdot\right)$ is defined for any age $j \in \left[1, \bar{j}\right]$ if $e = e_1$, only for $j \in \left[j\left(2\right) + 1, \bar{j}\right]$ if $e = e_2$ and only for $j \in \left[j\left(3\right) + 1, \bar{j}\right]$ if $e = e_3$.*

In the class of employment value functions special attention must be devoted to the value function of newly employed agents. By assumption 9 this conditional value is

$$\bar{J}\left(x_j, p_{j-1} = (1, a_j) \mid \mathrm{d}_j = 0\right) = \max_{e} \left\{W_j(\theta, e, z, a)\right\}_{e=e_1}^{e^*} \tag{2.14}$$

where e^* is the agent's own education level. It is evident that the conditional value of first-time employed equals the highest employment value among those available, and is therefore subject to (2.13).

It is possible to prove that the conditional value function of employment is monotonous, concave and smooth, and the optimal policy is single valued and continuous.

Proposition 6 *The conditional value function of any employment, $W_j\left(\cdot\right)$, has the following properties*

1. *it is monotonically increasing in (θ, z_j, a_j), concave and differentiable on $\Theta \times Z \times \bar{A}$;*
2. *the associated optimal policy $p_j = \left(a_{j+1}, \mathrm{d}_j = 0\right) \in \Gamma\left(x_j, p_{j-1}\right)$ is single-valued, and d_j and a_{j+1} are continuous functions in (θ, z_j, a_j).*

The next step is to examine the conditional value of education, that we call V_j.

Lemma 4 *The conditional value $V_j(\theta, e, z, a)$ for education participants at age $j \in \{1, .., j(3) - 1\}$ exists, is unique and is defined as*

$$\bar{J}(x_j, p_{j-1} \mid \mathrm{d}_j = 1) = V_j(\theta, e, z, a) = \max_{a'} \; u(c, f_e(\theta)) + \tag{2.15}$$
$$+ s_j\beta \int_Z \pi_{z'|z} \; \max\left\{V_{j+1}(\theta, e', z', a'), \left\{W_{j+1}(\theta, e, z', a')\right\}_{e=e_1}^{e^*}\right\} dz'$$

where e^* is the education level of the agent. Both V_j and W_j are subject to (2.3). Moreover, the conditional value for education participants at age $j(3)$, denoted as $V_{j(3)}(\theta, e, z, a)$, exists, is unique and is defined as

$$\bar{J}(x_j, p_{j-1} \mid \mathrm{d}_j = 1) = V_{j(3)}(\theta, e, z, a) =$$
$$= \max_{a'} \; u(c, f_e(\theta)) + s_j\beta \int_Z \pi_{z'|z} \; \max_e \left\{W_{j+1}(\theta, e, z', a')\right\}_{e=e_1}^{e_3} dz'$$

where W_{j+1} is the only integrand because further education is not an option at age $j > j(3)$.

Obviously V_j can be defined only for a specific subset of ages.

Remark 4 *By assumption 6, $V_j(\cdot)$ is defined only for $j \in [1, j(3)]$.*

Some interesting properties are associated to the value function of education.

Proposition 7 *The conditional value function of education, $V_j(\cdot)$, has the following properties:*

1. *it is monotonically increasing in (θ, z_j, a_j);*
2. *the associated optimal policy $p_j = (a_{j+1}, d_j = 1)$ is upper hemi-continuous in (θ, z_j, a_j);*
3. *it has both right and left derivative with respect to a_j. Moreover it is the case that $\frac{\partial V_j(x)}{\partial a_j}_+ \geq \frac{\partial V_j(x)}{\partial a_j}_-$;*

4. *the state space locations where* $\frac{\partial V_j(x)}{\partial a_j}_+ > \frac{\partial V_j(x)}{\partial a_j}_-$ *are never chosen. We refer to them as 'switch' points because at such locations a kink in* V_j *occurs, indicating a switch in the binary variable;*

5. *the associated optimal asset policy* a_{j+1} *is single-valued everywhere but at the 'switch' points .*

The fact that $\frac{\partial V_j(x)}{\partial a_j}_+ \geq \frac{\partial V_j(x)}{\partial a_j}_-$ implies that the education value is not generally concave, but only piece-wise concave. This is potentially troubling because it raises issues of non-uniqueness of the optimal policies and non-sufficiency of first order conditions, making the applicability of the Envelope Theorem problematic. However, since the kink in the upper envelope at the switch points is inward, it is never voluntarily chosen by an optimizing agent because, by continuity, there is a discrete payoff (marginal utility jump) by choosing another point arbitrarily close to the switch point. Moreover, given continuity of the assets domain, the probability of ending up in one of those switch points due to pure model randomness is arbitrarily close to zero. Therefore, having a tie-break condition is not necessary[13].

Remark 5 *The locations where switches in conditional values occur (switch points) correspond to inward kinks of* V_j *and are never chosen by optimising agents.*

The reason why the inward kink points are not chosen is that a marginal change in asset saving can guarantee a positive jump in marginal utility, which makes the kink point a suboptimal choice. We conclude that the optimal asset policy is unique at all relevant points and is piece-wise continuous between switch points.

These results characterize the unconditional choice problem of an agent who is still in education at age $j \leq j\,(3)$ as

$$\max_{\{a_{j+1},d_j\}} \{V_j, W_j\}$$

[13] A tie-break condition might still be necessary if the initial wealth endowments were non-continuously distributed.

We call this the unconditional problem because we are not restricting the value of d_j. A natural corollary follows.

Corollary 2 *The optimal unconditional policy $p_j = (a_{j+1}, d_j)$ is single-valued everywhere but at the 'switch' points of V_j. The only discontinuities of $p_j = (a_{j+1}, d_j)$ occur at the 'switch' points.*

The discontinuities in the asset policies occur at the switch points because of the jumps in marginal utility at such locations. Nonetheless, the optimal policy duplet $p_j = (a_{j+1}, d_j)$ is continuous between successive switch points.

2.5 Some properties of the optimal binary choice

In this section we discuss some features of the model and characterize a sufficient condition for the existence of a reservation policy for educational choice with respect to asset holding and ability. We start by arguing that for current asset holding a_j close enough to the lower bound $a_{\min}$ it is the case that $W_j(\theta, e, z, a) > V_j(\theta, e, z, a)$; this simply means that very poor people never go into education, regardless of their ability or idiosyncratic shock.

Remark 6 *There exists an $\varepsilon > 0$ such that for $a_{\min} < a \leq \varepsilon$ the following inequality $W_j(\theta, e, z, a) > V_j(\theta, e, z, a)$ holds for any triplet (θ, e, z).*

The previous remark follows from the definitions of the conditional value functions in (2.13-2.15), with the corresponding budget constraints, and the Inada conditions for utility.

A natural question is whether there exists an amount of asset holding that induces agents to choose education. We now provide a sufficient condition for the existence of at least one asset level at which education is chosen.

Proposition 8 *Let Assumption 2 hold, and $f^e(\cdot)$ be some (student leisure) function strictly increasing in ability θ, $0 \leq f^e(\theta) \leq 1$. Take a level of education and current* *shock,* (e, z), *and let* ε *be the defined as*

$$\varepsilon = \frac{Ra_j - a^*_{j+1} + \widetilde{w}n^*_j}{Ra_j - a^*_{j+1} - (D - Tr)} - 1$$

where $\frac{Ra_j - a^*_{j+1} + \widetilde{w}n^*_j}{Ra_j - a^*_{j+1} - (D - Tr)}$ *is the ratio of the consumptions when in education and employment for some given level of present asset holding* a_j *and where* $\left(a^*_{j+1}, n^*\right)$ *maximize the value of being employed. Given prices, tax rates and direct costs of education, and given some* c^* *maximising worker's utility, if there is a* θ *for which the following equality holds* [14]

$$u\left(\frac{c^*}{1+\varepsilon}, f^e(\theta)\right) = u(c^*, l^*)$$

and if $\frac{c^*}{1+\varepsilon}$ *is within the student's budget set, then* (θ, a) *are such that* $V_j(\theta, e_j, z_j, a^R_j) \geq W_j(\theta, e_j, z_j, a^R_j)$.

The above proposition provides a sufficient condition for education to be chosen over employment. It is important to notice that the above condition gives some insight on the relationship between current asset holding and educational choice: as the current asset holding a_j goes to $\frac{(D-Tr)}{R}$, the ε as defined above goes to infinity, implying that the equality $u\left(\frac{c^*}{1+\varepsilon}, f^e(\theta)\right) = u(c^*, l^*)$ cannot hold even for very high values of $f^e(\theta)$; similarly, as a_j becomes very large, ε tends to zero and at the same time an income effect makes l^* closer to one (agents supply little labor when they exit education) so that the equality described above can hold only for very high values of $f^e(\theta)$. In other words, people who are either very rich or very poor are less likely to remain in education for different reasons: very poor agents need to work in order to consume at the beginning of their lives, whereas very rich people choose to leave education because they can afford to consume large amounts of leisure when they are formally out of education.

Remark 7 *It is possible to show that for current asset holding* a_j *which are either*

[14] $u(\cdot, f^e(\theta))$ is the instantaneous utility when studying and $u(c^*, l^*)$ is the instantaneous utility when working, with the couple (c^*, l^*) maximizing the RHS of the equality. Of course the budget constraints are different for the LHS and the RHS.

sufficiently low or sufficiently high, the value of employment W_j lies above the value of education V_j .

It follows that if there exist a combination of (θ, a) such that the value of education is at least as big as the value of employment, then two cases are possible: (i) there is a finite subset of asset levels for which an agent always chooses education over employment, or (ii) the upper envelope of the two conditional value functions corresponds to the conditional value of employment at all levels of current assets.

Finally we provide a parametric equivalent of the condition spelled out in Proposition 8 for one kind of preferences that we will use in other chapters, specifically CRRA in leisure and consumption.

Corollary 3 *Let the utility function be of the CRRA kind described in (2.16-2.17) in the appendix with a given parameter ν and let $f^e(\theta)$ be some (student leisure) function which is strictly increasing in ability θ, $0 \leq f^e(\theta) \leq 1$. Let $\vec{w}$ be a finite vector of market-clearing wages in the economy, (τ_{n^e}, τ_k) be the current tax rates on labour income and asset returns, and D_e and Tr_e be, respectively, the direct monetary cost of and subsidy to education. Take any duplet (e_j, z_j), some current asset holding $\bar{a}_j$ and optimal asset saving when employed $a^*_{j+1}(x_j)$. Define the ratio $(1+\varepsilon)$ as follows*[15]

$$\frac{R\bar{a}_j - a^*_{j+1} + \widetilde{w}}{R\bar{a}_j - a^*_{j+1} - (D_e - Tr_e)} = 1 + \varepsilon$$

for an a^* maximising the value of employment which is also within the student's budget set. If there exists a θ such that $f^e(\theta) = (1+\varepsilon)^{\frac{\nu}{1-\nu}} l^*$, then there exists (at least one) current asset level a_j for which $V_j(\theta^R, e_j, z_j, a^R_j) \geq W_j(\theta^R, e_j, z_j, a^R_j)$ holds.

We incidentally notice that for $0 < \nu < 1$ we have that $f^e(\theta)$ must be larger than the optimal leisure l^*.[16].

[15] Here $\hat{w} = w_e \exp^{\epsilon^e}(1 - \tau_{n^e})$ and $R = (1 - \tau_k)(1 + r)$.

[16] Some useful analytical results for the CRRA utility are discussed in the Appendix 2 and will be used in the following chapters

2.6 Conclusions

In this chapter we examine how the inclusion of non-convexities in the choice set due to binary decision variables might affect the solution of an otherwise standard life-cycle model of consumption with uninsurable idiosyncratic risk. Simple conditions necessary for the existence of a functional equation representation of the solution are presented. We investigate how the standard properties of optimal policies change because of the binary decision, and characterize the behaviour of the binary decision in some critical state-space locations.

We find that under very standard assumptions a solution of the problem exists and can be described through functional equations. We also find that the optimal policies are always upper hemicontinuous but not single-valued at all points of the state-space. It follows that continuity of the optimal policies is lost through the inclusion of binary choices.

However, we argue that for the purposes of numerical implementation the optimal policies can be treated as if they were single valued, because the state-space locations where they are not single-valued are never chosen and, in presence of a continuous shock process, have a probability of occurring arbitrarily close to zero.

Of course, the lack of continuity of the optimal policies represents a very undesirable feature for numerical work, which cannot rely on simple local methods to identify optimal behaviour.

We also provide some results describing the interaction between discrete choices and heterogeneity. In particular, we show how the inclusion of different types of heterogeneity makes for more interesting and realistic dynamic binary choices.

2.7 Appendix to chapter 2

2.7.1 First order conditions and budget sets: education and labour supply decisions with CRRA utility

The period utility function for an agent in full time education is

$$u(c) = \frac{\left[c^{\nu} f^{e}(\theta)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda} \tag{2.16}$$

where $f^{e}(\theta)$ is a monotonically increasing function of the innate ability parameter θ.

The period utility for an employed agent is instead

$$u(c,l) = u(c,1-n) = \frac{\left[c^{\nu}(1-n)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda} \tag{2.17}$$

The analytical forms of $u_c(c,l)$ and $u_n(c,l)$ for this period utility of an employed agent are

$$u_c(c,l) = \left(c^{\nu}(1-n)^{1-\nu}\right)^{-\lambda} \nu c^{\nu-1}(1-n)^{1-\nu} \tag{2.18}$$

$$u_l(c,l) = \left(c^{\nu} l^{1-\nu}\right)^{-\lambda}(1-\nu) c^{\nu} l^{-\nu}$$

so that $\frac{u_l(c,l)}{u_c(c,l)} = \left(\frac{1-\nu}{\nu}\right)\left(\frac{c}{l}\right)$.

From the intra-temporal margin we know that $\left(\frac{1-\nu}{\nu}\right)\frac{c}{l} = w_e \exp^{\epsilon^e}(1-\tau_n)$ and solving this equality for $n = 1-l$ we get the optimal supply of labor as a function of consumption

$$n = \max\left\{1 - \left(\frac{1-\nu}{\nu}\right)\frac{c}{w_e \exp^{\epsilon^e}(1-\tau_{n^e})}, 0\right\} \tag{2.19}$$

If we plug equation (2.19) into the period budget constraint of a working agent,(2.3), we can cancel out labor supply and obtain a *'consumption only'* budget constraint

$$c_j + a_{j+1} = \left[1 + r(1-\tau_k)\right] a_j + w_e \exp^{\epsilon_j}(1-\tau_{n^e})\left(1 - \frac{1-\nu}{\nu}\frac{c_j}{w_e \exp^{\epsilon_j}(1-\tau_{n^e})}\right)$$

$$\Longrightarrow \nu(c_j + a_{j+1}) = \nu\left[1 + r(1-\tau_k)\right] a_j + \nu w_e \exp^{\epsilon_j}(1-\tau_{n^e}) - (1-\nu) c_j$$

$$\Longrightarrow c_j + \nu a_{j+1} = \nu\left[1 + r(1-\tau_k)\right] a_j + \nu w_e \exp^{\epsilon_j}(1-\tau_{n^e}) \tag{2.20}$$

Finally, by using the intra-temporal margin, we can express the period utility as a function of consumption only

$$\frac{\left[c^{\nu}\left(\frac{1-\nu}{\nu}\frac{c}{w_e \exp^{\epsilon}(1-\tau_{n_e})}\right)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda} = \frac{\left[\left(\frac{1-\nu}{\nu}\frac{1}{w_e \exp^{\epsilon}(1-\tau_{n_e})}\right)^{1-\nu} c\right]^{(1-\lambda)}}{1-\lambda} \tag{2.21}$$

Furthermore we can derive an analytical solution for the labor supply function $n_j = n_j(\theta, e, z, a)$, given w_e. For notational simplicity we write eq.(2.20) as $c_j = \nu R a_j + \nu\widetilde{w} - \nu a_{j+1}$, where $R = [1 + r(1-\tau_k)]$ and $\widetilde{w} = w_e \exp^{\epsilon_j}(1-\tau_{n^e})$. Then the optimal labor supply function is given by

$$\begin{aligned} n_j &= \max\left\{1 - \left(\frac{1-\nu}{\nu}\right)\frac{\nu R a_j + \nu\widetilde{w} - \nu a_{j+1}}{\widetilde{w}}, 0\right\} \\ &= \max\left\{\nu + (1-\nu)\frac{a_{j+1} - R a_j}{\widetilde{w}}, 0\right\} \end{aligned} \tag{2.22}$$

This expression is useful to analyze the life-long pattern of labor supply and is nothing else than a weighted average of 1 and $\frac{a_{j+1}-Ra_j}{\widetilde{w}}$, with weights equal to ν and $(1-\nu)$.

ν is the fraction of labor supply directly related to providing utility through consumption, whereas $(1-\nu)$ is the leisure-related component of labor supply, depending on income and substitution effects.

Notice that $\frac{a_{j+1}-Ra_j}{\widetilde{w}} \leq 1$, if the budget constraint holds. If $\frac{a_{j+1}-Ra_j}{\widetilde{w}} = 1$ it follows that $n_j = 1$; if $a_{j+1} = Ra_j$ then $n_j = \nu$. Finally, when $\frac{a_{j+1}-Ra_j}{\widetilde{w}} \leq \frac{-\nu}{1-\nu}$ we have that $n_j = 0$.

This simple relationship can go quite far in explaining the labor supply profile of an agent with finite life as agents accumulate assets at the beginning of their life (that is, when $a_{j+1} > Ra_j$) we can expect relatively high labor supply, whereas at later stages in life, when agents deplete their asset stock, labor supply decreases and, if $\widetilde{w} = w_e \exp^{\epsilon_j}(1-\tau_{n^e})$ is small enough, it can get arbitrarily close to zero.

2.7.2 Proofs of lemmas and propositions

Proof of Lemma 1. At each j, $a_{\bar{j}+1}$ is bounded as shown in equation (2.4). Say that $\bar{a}_{j+1} < \infty$ is the upper bound of a_{j+1}, then we can define the set

$$B(\bar{x}_1) \equiv \prod_{j=1}^{\bar{j}} [0, \bar{a}_{j+1}]$$

The feasible set $B(\bar{x}_1)$ is the finite product of compact sets and it follows that $B(\bar{x}_1)$ is compact in the $\bar{j}$ product topology. This result is also known as Tychonoff theorem[17].

■

Proof of Proposition 1.

We can find the upper-bound of $U(\bar{x}_1, c, l)$ subject to the resource constraint (2.3). From (2.6) we know that consumption has an upper bound at any j, $c_j \le \bar{c}_j < \infty$. Using the intra-temporal margin we can express $U(\bar{x}_1, c, l)$ as a function of consumption only $U(\bar{x}_1, c)$ and given $u(\bar{c}_j) < \infty$ we can write

$$U(\bar{x}_1, c) = E\sum_{j=1}^{\bar{j}} \left(s_j \beta^{j-1}\right) u(c_j) \le E\sum_{j=1}^{\bar{j}} \left(s_j \beta^{j-1}\right) u(\bar{c}_j) < \infty$$

■

Proof of Proposition 2. Given initial conditions $\bar{x}_1 \in \Delta$, if the feasible set $B(\bar{x}_1)$ is compact, and the indirect utility $\breve{U}(\bar{x}_1, p)$ is upper hemi-continuous on $B(\bar{x}_1)$, then, by the Weierstrass Theorem, an optimal plan exists. $B(\bar{x}_1)$ is compact by Lemma 1.

To prove the upper hemi-continuity requirement, we first show that $c(p)$ is continuous in p, and that $U(\bar{x}_1, c, l)$ is upper hemi-continuous in c. Then, it follows that $\breve{U}(\bar{x}_1, p)$ is upper hemi-continuous on $p \in B(\bar{x}_1)$[18].

In order to show continuity of $c(p) : B(\bar{x}_1) \to \Re_+$ in $p \in B(\bar{x}_1)$, we need to show that all its elements $c_j(p_j)$ are continuous in p_j. In equation (2.6) c_j is defined as a

[17]See Becker and Boyd (1997), page 41.
[18]This result is from Stokey, Lucas, and Prescott (1989), page 58.

function of $p_j = (\mathrm{d}_j(x_j), a_{j+1}(x_j))$ where $x_j = (\theta, e_j, z_j, a_j)$. The function $c_j(p_j)$ is continuous in (d_j, a_{j+1}) if and only if $\left(\mathrm{d}_j^i, a_{j+1}^i\right) \rightarrow (\mathrm{d}_j, a_{j+1})$ implies $c_j\left(p_j^i\right) \rightarrow c_j(p_j)$. Note that a convergent sequence in $\{0, 1\}$ must be zeros before converging to zero, and ones before converging to one using this result it is easy to check that, given some $\mathrm{d}_j^i \rightarrow \mathrm{d}_j$, c_j as defined in equation (2.6) satisfies the above condition as $a_{j+1}^i \rightarrow a_{j+1}$. So c_j is continuous in p_j.

Finally, we have to show that $U(\bar{x}_1, c, l) : \Re_+ \times [0,1] \rightarrow \Re$ is upper hemi-continuous on $c \in \Re_+$.[19] Conditional on the period shock z_j, take some $\hat{c} \in \Re^{\bar{j}}$ and a sequence $c^i \in \Re^{\bar{j}}$ such that $c^i \rightarrow \hat{c}$. Then

$$\lim_{i\rightarrow\infty} \sup U(\bar{x}_1, c) = \lim_{i\rightarrow\infty} \sup \sum_{j=1}^{\bar{j}} s_j \beta^{j-1} u\left(c_j^i(z_j)\right) \pi\{z_j \mid z_{j-1}\}$$

Note that, from equation (2.6) the consumption sequence c^i has an upper bound at every j that we denote as $\bar{c}_j < \infty$,

$$c_j^i \leq \bar{c}_j \quad \forall j$$

Then the instantaneous utility sequence has also a finite upper bound

$$s_j \beta^{j-1} u\left(c_j^i\right) \leq s_j \beta^{j-1} u(\bar{c}_j) < \infty \quad \forall j$$

By the Inada condition $u(\bar{c}_j) > -\infty$; then

$$-\infty < \sum_{j=1}^{\bar{j}} s_j \beta^{j-1} u\left(c_j^i(z_j)\right) \pi\{z_j \mid z_{j-1}\} < \infty$$

[19] We do not concern ourselves with leisure l because the intra-temporal condition for leisure allows to reduce the instantaneous utility to a function of c_j alone.

and we can write the following inequality[20]

$$\lim_{i\to\infty}\sup\sum_{j=1}^{\bar{j}} s_j\beta^{j-1}u\left(c_j^i\left(z_j\right)\right)\pi\{z_j \mid z_{j-1}\} \leq \sum_{j=1}^{\bar{j}} \lim_{i\to\infty}\sup s_j\beta^{j-1}u\left(c_j^i\left(z_j\right)\right)\pi\{z_j \mid z_{j-1}\}$$

Finally we use the upper hemi-continuity of the instantaneous utility $u(\cdot)$ in c to write

$$\lim_{i\to\infty}\sup s_j\beta^{j-1}u\left(c_j^i\left(z_j\right)\right) \leq s_j\beta^{j-1}u\left(\hat{c}_j\left(z_j\right)\right)$$

for every j,[21] so that

$$\sum_{j=1}^{\bar{j}} \lim_{i\to\infty}\sup s_j\beta^{j-1}u\left(c_j^i\left(z_j\right)\right)\pi\{z_j \mid z_{j-1}\} \leq \sum_{j=1}^{\bar{j}} s_j\beta^{j-1}u\left(\hat{c}_j\left(z_j\right)\right)\pi\{z_j \mid z_{j-1}\}$$

and hence

$$\lim_{i\to\infty}\sup U\left(\bar{x}_1, c^i\right) \leq U\left(\bar{x}_1, \hat{c}\right)$$

which is a sufficient condition for the upper hemi-continuity of $U(\bar{x}_1, c)$ on $c \in \Re_+$. ∎

Proof of Lemma 2. Since $\breve{U}(\bar{x}_1, p)$ is upper hemi-continuous in $(\bar{x}_1, p)$, and since Γ is a continuous mapping, then $U^*(\bar{x}_1)$ is non-empty-valued, upper hemi-continuous in $(\bar{x}_1)$ and compact-valued by Berge's maximum theorem. See also Townsend and Ueda

[20] To prove the inequality just take a sequence of J-periods plans conditional on z_j, $g^m = \sup\{f^m, f^{m+1}, ...\}$, so that $g^m \geq f^n$ for $n \geq m$. Then, given some initial condition $\bar{z}_1$, $\sum_{j=1}^{J} g_j^m\pi\{z_j \mid z_{j-1}\} \geq \sum_{j=1}^{J} f_j^n\pi\{z_j \mid z_{j-1}\}$. It follows that

$$\sum_{j=1}^{J} g_j^m\pi\{z_j \mid z_{j-1}\} \geq \lim_{n\to\infty}\sup\sum_{j=1}^{J} f_j^n\pi\{z_j \mid z_{j-1}\}$$

Finally, note that

$$\sum_{j=1}^{J} \lim_{n\to\infty}\sup f_j^n\pi\{z_j \mid z_{j-1}\} \geq \sum_{j=1}^{J} g_j^m\pi\{z_j \mid z_{j-1}\}$$

by definition of lim sup, which delivers the result

$$\sum_{j=1}^{J} \lim_{n\to\infty}\sup f_j^n\pi\{z_j \mid z_{j-1}\} \geq \lim_{n\to\infty}\sup\sum_{j=1}^{J} f_j^n\pi\{z_j \mid z_{j-1}\}$$

[21] In this case we are using the fact that $u(\cdot)$ is by assumption continuous in its arguments and we are applying proposition 2, page 47 of Becker and Boyd (1997), claiming that a function f is u.h.c. if and only if $f(x^*) \geq \limsup_{i\to\infty} f\left(x^i\right)$ whenever $x_i \to x^*$.

(2003) and Berge (1997), page 116. Upper hemi-continuity and compact-valuedness of the correspondence $U^*(\bar{x}_1)$ imply that for any $\bar{x}_1^i \to \bar{x}_1^*$ there exists a convergent subsequence of $\{U^*(\bar{x}_1^i)\}$ whose limit, as $i \to \infty$, is in $U^*(\bar{x}_1^*)$. Since the supremum of any function is unique, $U^*(\bar{x}_1)$ is single-valued, and therefore the sequential characterisation above implies continuity of $U^*(\bar{x}_1)$.

Given that the supremum of measurable functions is also measurable[22], then $U^*(\bar{x}_1)$ is measurable by assumption 8.

As for monotonicity, assuming $U^*(\bar{x}_1 = \bar{\theta}, e_1, \bar{z}_1, \bar{a}_1)$ is achieved by an optimal sequence p^*, this optimal sequence is also feasible for any other $x_1 = (\theta, e_1, z_1, a_1)$ such that either $\theta > \bar{\theta}$ or $z_1 > \bar{z}_1$ or $a_1 > \bar{a}_1$, or any combination of the three. ■

Proof of Proposition 3. By definition in (2.11)

$$J(x_1) = \sup_{p_1 \in \Gamma(x_1, p_0)} \mathrm{v}(p_1, \bar{x}_1, p_0) + (s_1\beta) \int_Z \pi(z_2 \mid z_1) J(x_2, p_1)\, dz_2$$

in this expression we can replace $J(x_2, p_1)$ with its definition and use the definition of v in (2.7) to obtain

$$\begin{aligned} &= \sup_{p_1 \in \Gamma(x_1, p_0)} \mathrm{v}(p_1, \bar{x}_1, p_0) + \\ &+ (s_1\beta) \int_Z \pi(z_2 \mid z_1) \left[\sup_{p_2 \in \Gamma(x_2, p_1)} \mathrm{v}(p_2, x_2, p_1) + (s_2\beta) \int_Z \pi(z_3 \mid z_2) J(x_3, p_1)\, dz_3 \right] dz_2 \end{aligned}$$

then, following Townsend and Ueda (2003), we can write

$$\begin{aligned} &= \sup_{p_1 \in \Gamma(x_1, p_0), p_2 \in \Gamma(x_2, p_1)} \mathrm{v}(p_1, \bar{x}_1, p_0) + \\ &+ (s_1\beta) \int_Z \pi(z_2 \mid z_1) \left[\mathrm{v}(p_2, x_2, p_1) + (s_2\beta) \int_Z \pi(z_3 \mid z_2) J(x_3, p_1)\, dz_3 \right] dz_2 \end{aligned}$$

[22] See Stinchcombe and White (1992).

which in turn gives

$$= \sup_{p_1\in\Gamma(x_1,p_0),p_2\in\Gamma(x_2,p_1)} \mathrm{v}\left(p_1,\bar{x}_1,p_0\right)+\left(s_1\beta\right)\int_Z \pi\left(z_2\mid z_1\right)\mathrm{v}\left(p_2,x_2,p_1\right)dz_2+$$
$$+\left(s_1s_2\beta^2\right)\int_Z \pi\left(z_2\mid z_1\right)\pi\left(z_3\mid z_2\right)J\left(x_3,p_1\right)dz_3dz_2$$

By applying this substitution $\bar{j}$ times, we get

$$J\left(x_1\right)=\sup_{\{p_j\in\Gamma(x_j,p_{j-1})\}_{j=1}^{\bar{j}}} E_Z\sum_{j=1}^{\bar{j}}\left(\prod_{i=1}^{j}s_i\right)\beta^{j-1}\mathrm{v}\left(p_j,x_j,p_{j-1}\right)$$

which coincides with the sequence problem in (2.9). ■

Proof of Proposition 4. See Stokey, Lucas, and Prescott (1989), theorem 9.4, page 251. For a discussion of how the necessary assumptions are satisfied, see Townsend and Ueda (2003). ■

Proof of Proposition 5. See Stokey, Lucas, and Prescott (1989), theorem 9.6, page 263. Assumptions 9.4-9.7 are satisfied, as Z is compact by assumption 1 and its transition function $\pi\left(\cdot\right)$ has the Feller property, by assumption 5, Γ is non-empty, compact-valued and continuous, the utility is bounded. ■

Proof of Lemma 3. For the existence and the uniqueness of the conditional value function see Proof of Proposition 5. All necessary assumptions are trivially satisfied. As for the definition of the employment value, it follows from Assumptions 9,6,4 and 5. ■

Proof of Proposition 6. Monotonicity of $W_j(\theta,e,z,a)$ follows from the same arguments used in the proof of Lemma 2. Concavity of $W_j(\theta,e,z,a)$ can be proved by backward induction by assumption 3 and the demographic fact that $\left(\left(1-s_{\bar{j}}\right)=1\right.$, $W_{\bar{j}}(\theta,e,z,a)=u\left(c\left(\theta,e,z,a\right),n\left(\theta,e,z,a\right)\right)$, which is strictly concave in both its arguments by assumption 2. By assumption 4, if $W_{\bar{j}}(\theta,e,z,a)$ is strictly concave, then such is $\int_Z \pi_{z'|z} W_{\overline{j}}(\theta, e, z', a')\, dz'$

$$W_{\overline{j}-1}(\theta, e, z, a) = \max_{a', n} u\,(c, 1-n) + s_j\beta \int_Z \pi_{z'|z}\, W_{\overline{j}}\left(\theta, e, z', a'\right)\, dz'$$

is also concave, as it is the sum of two concave functions. We can repeat this argument going backwards up to the first age when W_j can be defined. Differentiability of $W_j(\theta, e, z, a)$ follows from concavity, see Stokey, Lucas, and Prescott (1989), Theorem 4.11, page 85. The conditions for the theorem are satisfied by Assumptions 2 and 1 and Propositions 5 and 1. Furthermore, the necessary Assumption 4.8 in Stokey, Lucas, and Prescott (1989), page 80, is satisfied by the fact that, by assumption 5, when $\mathrm{d}_j = 0$ the set of feasible policies is reduced to

$$p_j \in \Gamma\left(x_j, \mathrm{d}_{j-1}\right) \equiv 0 \times \left[0, Ra_j + \widetilde{w}n_j\right]$$

and Γ is convex in the sense of Stokey, Lucas, and Prescott (1989). In order to prove that $p_j \in \Gamma\left(x_j, \mathrm{d}_{j-1}\right)$ is single-valued and that d_j and a_{j+1} are continuous, we first resort to Proposition 5 (Berge's Maximum Theorem) to claim that the optimal policies correspondence (2.12) is non-empty, compact-valued and upper hemi-continuous. Then, we notice that by Assumption 5 the value of d_j is going to be zero thereafter; finally, using the strict concavity of $u\,(\cdot)$ from Assumption 2, and noticing that when $\mathrm{d}_j = 0$ the feasible set $\Gamma\left(x_j, \mathrm{d}_{j-1}\right) \equiv 0 \times \left[0, Ra_j + \widetilde{w}n_j\right]$ is convex-valued, we argue that the optimal asset policy $a_{j+1}\,(x)$ is indeed unique. Since the optimal policies correspondence (2.12) is upper hemi-continuous, d_j is a sequence of zeros and the asset policy is single valued, we conclude that d_j and a_{j+1} are also continuous functions (see also the proof of Lemma 2 for a discussion of this point). ■

Proof of Lemma 4. Same as Proof of Lemma 3. ■

Proof of Proposition 7. Monotonicity of $V_j(\theta, e, z, a)$ follows from the same arguments used in the proof of Lemma 2.

In order to prove that the optimal policy $p_j = (a_{j+1}, d_j = 1)$ is upper hemi-continuous in (θ, z_j, a_j) we use Proposition 5 (Berge's Maximum Theorem) to claim that the optimal policies correspondence (2.12) is non-empty, compact-valued and upper hemi-continuous.

The existence of left and right derivatives and the property $\frac{\partial V_j(x)}{\partial a_{j+1}}_+ \geq \frac{\partial V_j(x)}{\partial a_{j+1}}_-$ can be proved as follows. First, resort to the definition of education value

$$V_j(\theta, e, z, a) = \max_{a_{j+1}} u\left(c_j, f_e(\theta)\right) +$$
$$+ s_j\beta \int_Z \pi_{z_{j+1}|z_j} \max\left\{V_{j+1}(\theta, e_{j+1}, z_{j+1}, a_{j+1}), \left\{W_{j+1}\left(\theta, e_{j+1}, z_{j+1}, a_{j+1}\right)\right\}_{e_{j+1}=e_1}^{e_{max}}\right\} dz_{j+1}$$

The first component $u\left(c_j, f_e(\theta)\right)$ can be differentiated with respect to a_j. The second component is the upper envelope of future realizations of V_{j+1} and W_{j+1}. We know that W_{j+1} is concave in a_j and therefore differentiable. Moreover, V_{j+1} is the upper envelope of future values of education and/or work. The upper envelope of concave and piece-wise concave functions is itself piece-wise concave. Then, given some (θ, e, z), V_j can be described as a succession of montonous and concave functions over i asset intervals $\left(\underline{a}_j^{(i)}, \overline{a}_j^{(i)}\right)$ such that

- for any $a_j \in \left(\underline{a}_j^{(i)}, \overline{a}_j^{(i)}\right)$, both $\frac{\partial V_j(x)}{\partial a_j}_+$ and $\frac{\partial V_j(x)}{\partial a_j}_-$ exist, and $\frac{\partial V_j(x)}{\partial a_j}_+ = \frac{\partial V_j(x)}{\partial a_j}_-$;
- for $a_j = \underline{a}_j^{(i)}$, $\frac{\partial V_j(x)}{\partial a_j}_+$ exists ;
- for $a_j = \overline{a}_j^{(i)}$, $\frac{\partial V_j(x)}{\partial a_j}_-$ exists ;
- and $\frac{\partial V_j(x)}{\partial a_j}_{+|a_j=\underline{a}_j^{(i)}} \geq \frac{\partial V_j(x)}{\partial a_j}_{-|a_j=\overline{a}_j^{(i-1)}}$.

Then the boundary points of each interval $\left(\underline{a}_j^{(i)}, \overline{a}_j^{(i)}\right)$ are those where the ends of two consecutive concave functions meet. At such points it must be the case that $\frac{\partial V_j(x)}{\partial a_j}_+ > \frac{\partial V_j(x)}{\partial a_j}_-$. As a proof, suppose that there exists some boundary point $a_j^{(i)}$ where $\frac{\partial V_j(x)}{\partial a_j}_+ \leq \frac{\partial V_j(x)}{\partial a_j}_-$, then it would not be optimal to switch conditional value at $a_j^{(i)}$, and therefore $a_j^{(i)}$ would not be a boundary point which contraddicts the initial statement.

Finally, it is possible to prove that a_j is single-valued everywhere but at the switch points $\frac{\partial V_j(x)}{\partial a_{j+1}}_+ > \frac{\partial V_j(x)}{\partial a_{j+1}}_-$ using Proposition 5 (Berge's Maximum Theorem) to claim that the optimal policies correspondence (2.12) is non-empty, compact-valued and upper hemi-continuous. Then, we notice that the value of d_j is constant over a given concave interval, so that the feasible set $\Gamma(x_j, \mathrm{d}_{j-1})$ is convex-valued, and uniqueness of the optimal asset policy a_{j+1} follows. ■

Proof of Proposition 8. First consider the ratio of the consumptions when in education and employment for some given level of present asset holding a_j, $\frac{Ra_j - a_{j+1} + \widetilde{w}n_j}{Ra_j - a_{j+1} - (D - Tr)}$, and then define ε as

$$\varepsilon = \frac{Ra_j - a^*_{j+1} + \widetilde{w}n^*_j}{Ra_j - a^*_{j+1} - (D - Tr)} - 1$$

where a^*_{j+1} denote the optimal asset saving when employed and n^*_j the optimal labor supply. Given some (e, z), if there exists a θ such that $u\left(\frac{c^*}{1+\varepsilon}, f^e(\theta)\right) = u(c^*, l^*)$ for an ε defined as above and $\frac{c^*}{1+\varepsilon}$ within the student's budget set, with (c^*, l^*) maximizing W_j, then the current asset holding a_j that can achieve such ε is such that $V_j \geq W_j$. This last inequality follows because, keeping present utility constant, the optimal policy a^*_{j+1} and the associated employment value for next period is one of the available options when in education, which means that the value of education is at least as big as the value of employment. ■

Chapter 3

Education Decisions, Equilibrium Policies and Wage Dispersion

3.1 Introduction

This chapter examines the effects of alternative policies on the distribution of education in both partial and general equilibrium. Empirical evidence suggests a link between human capital accumulation and wages dispersion (see for example Mincer (1991, 1994)), so that policies affecting education outcomes will also have an impact on inequality, productivity and welfare. We use a life-cycle model of labor earnings with endogenous labour supply and education choice, allowing for agents' heterogeneity in several dimensions.

Individual choices are analyzed in the context of a general equilibrium model with separate, education-specific spot markets for jobs. The unit price of (efficiency-weighted) labour differs by education group and equals marginal product.

We are interested in the equilibrium, long-term effects of policy interventions targeting the wider population rather than limited groups, with relative labour prices endogenously adjusting to changes in the aggregate supply of educated people[1].

[1]Admittedly, given that labor is bought and sold on spot markets,the demand for labor is always equal to the supply. Alternatively, Acemoglu (2002) studies a model in which the demand for skills changes more than proportionally as a response to the increase in the supply of skilled workers.

We examine traditional policies, such as tuition transfers and subsidies[2], but our structure could be used to evaluate a number of alternative forms of policy intervention. The policy experiments are carried out through numerical simulations, with some of the model's parameters directly estimated from PSID and CPS data and others calibrated to match specific long-term features of the US economy. By simulating and comparing equilibrium outcomes we aim to explore the quantitative aspects of the relationship among education participation, endogenous selection, wages inequality and education policy.

When we experiment with college tuition subsidies it becomes apparent that while in partial equilibrium such policies can be very effective in increasing education levels and reducing inequality in general equilibrium the results are much less encouraging: the main effect of a subsidy there is to increase the supply of human capital as one would expect. However, it is the more able but liquidity constrained individuals who take up extra education, while the education levels of the less able can actually decrease (they are crowded out). Thus the subsidy acts on the composition of those in education. In many respects this is in line with results found by Heckman, Lochner, and Taber (1998b). The inclusion of risky returns on labor earnings and the fact that labor supply is endogenous lend additional credibility to the result.

3.1.1 Literature review

Research linking human capital investment to life cycle earnings dates back to original work by Mincer (1958), Becker (1964) and Ben-Porath (1967). The first studies ignored the important issue of self selection into education, as described by Rosen (1977) and Willis and Rosen (1979). Permanent and transitory individual characteristics are now acknowledged as important determinants of education choices and have become a standard feature of HC models. Empirical evidence supporting the plausibility of a link between human capital accumulation and economic inequality has been provided, among others, by Mincer (1991).

[2]Standard education policy is just one of the possible types of human capital policy. For example, changes in proportional income taxation affect the life-cycle returns on HC and the opportunity costs of education, altering HC investment decisions.

In work relating education policies and individual preferences Fernandez and Rogerson (1995) originally point out that heterogeneity among individuals, whether in terms of income, ability or locality, can generate conflicting preferences as to the kind of policies that are most desirable[3].

Studies on the evaluation of policy interventions in equilibrium are more recent. Heckman, Lochner, and Taber (1998b, 1998c) have led the way in advocating an approach to policy evaluation which does not overlook equilibrium effects induced by the policy[4]. In fact, statements regarding the effects of policy interventions which ignore price changes induced by such interventions are misleading. Fernandez and Rogerson (1998) provide an interesting application of G.E. modelling to the evaluation of education-finance reform in the US. Later work by Cunha, Heckman, and Navarro (2004, 2005) reinforces the view that models that are able to construct equilibrium counterfactuals are essential to understanding the wider consequences of policy interventions.

In the empirical literature on education policy, early work by Keane and Wolpin (1997) focuses on the partial equilibrium effect of a tuition subsidy on young males' college participation. A valuable generalization of their approach within a dynamic GE framework is due to Lee (2001). Also Abraham (2001) examines wage inequality and education policy in a GE model of skill biased technological change. All these studies restrict labor supply to be fixed, although earlier theoretical research has uncovered interesting aspects of the joint determination of life cycle labor supply and HC investment, among others Blinder and Weiss (1976).

Our model incorporates two twists with respect to earlier work: first, optimal individual labor supplies are an essential part of the lifetime earnings mechanism; second, agents' heterogeneity has different dimensions, including a permanent (ability) component and a persistent efficiency shock [5]. Each agent in our model represents a single individual household, consistently with the empirical analysis we produce.

Recent empirical evidence in Hyslop (2001) indicates that labor supply explains

[3]Fernandez and Rogerson (1995) consider ex-ante identical individuals who differ only in income

[4]Heckman, Lochner, and Taber estimate and simulate a dynamic general equilibrium model of education accumulation, assets accumulation and labor earnings with skill-biased technological change.

[5]Mortality risk is also explicitly included in the model.

little of the rising earnings inequality for married men, but over 20% of the rise in (both permanent and transitory) family inequality during the period of rising wage inequality in the early 1980's [6]. Moreover, the response in hours of work to changes in net wage is small for prime age male earners. However, as pointed out by Eaton and Rosen (1980) in their seminal work on taxation and HC accumulation, even if taxes have only a limited impact upon the *quantity* of hours worked it is possible that they have an important effect on their *quality*, intended as the type of human capital. This happens because tax changes can alter the incentives for education. Moreover, even if individual labor supplies do not deviate much from some average levels, it is the case that average levels may differ between education groups. For given market prices, work effort represents the intensity of human capital utilization and individuals can self-select into education groups according to their preference for leisure [7]. Labor supply, therefore, represents an effective channel of adjustment to labor price signals and an important determinant of the relative variation in skill prices[8].

The other crucial twist in our model is the introduction of individual uncertainty over the returns to HC in the form of idiosyncratic multiplicative shocks to labor efficiency. As Levhari and Weiss (1974) originally emphasized, uncertainty is of exceptional importance in human capital investment decisions as the risk associated to such decisions is usually not insurable nor diversifiable. Problems of moral hazard can be extremely severe when insuring labor risk because idiosyncratic shocks and individual ability can be partially or completely unobservable to third parties. Given these problems the market is not likely to provide insurance. Using a multiplicative form of earnings risk[9] Eaton and Rosen (1980) show how earnings taxation has an ambiguous effect on investment in human capital because it impinges on two important parameters of the decision problem: for one, taxation reduces the riskiness of returns

[6] Hyslop (2001) also shows that labor supply explains roughly half of the modest rise in female inequality.

[7] This selection in our model happens through permanent unobserved characteristics.

[8] Consider, for example, taxes on labor earnings which reduce the return to HC investment but also the opportunity cost of being in education represented by foregone earnings. When differences in lifetime labor supply between education groups are present, the two effects are weighted by the relative intensity of HC utilization in the appropriate education group.

[9] They multiply education specific earnings by a random variable.

to human capital investment;[10] in addition, taxation induces an income effect that can influence the agents' willingness to bear risk. Thus, ignoring the riskiness of education decisions can partly sway the results in the analysis of the effects of earnings taxation and education policies.

3.2 Model

This section describes the model we use for the analysis of education policy. In many respects this model is similar to the one described in chapter 2. The main difference lies in the introduction of an explicit production technology which delivers the marginal productivity of various inputs.

We model three levels of education obtained through formal schooling and corresponding to three types of human capital which enter the production technology.[11] Education and employment are mutually exclusive in each period. Foregone earnings and tuition charges are the direct costs of schooling, and an additional utility cost is associated to schooling through reductions in leisure.[12]

Agents can accumulate assets and we assume that the distribution of assets among newborns is the same as the distribution of accidental bequests in the economy.[13] The structure of our model allows to introduce different levels of correlation between ability and initial assets holdings.[14]

In general, the model allows us to look at endogenous equilibrium levels of aggregate human capital, with associated wages, as a function of agents' optimizing schooling

[10]As the proportional tax rate increases, agents earn less from high realization of the shock but also lose less from the bad ones. Therefore the overall risk is decreased.

[11]We distinguish among people with less than high school degrees (LTHS), high school graduates (HSG) and college graduates (CG). The distinction between LTHS and HSG is based on different earning and labor supply characteristics. Schooling is the only way to accumulate human capital (no children nurturing or on-the-job training). The possible effects of OJT are accounted for through an age-efficiency profile which is estimated for each education group and is maintained to be policy-invariant.

[12]These utility costs are calibrated to match enrolment rates for different ability groups

[13]Gale and Scholz (1994) show that inter vivos transfer for education represent only a part of total bequest. We ignore this issue in this paper and redistribute assets only among the youngest. The initial distribution of wealth replicates the distribution prevailing among those who die in each period, as this has a realistic accidental bequest interpretation.

[14]This can be thought of as a shortcut to incorporate the effect of parental background on ability formation, as extensively documented in the literature, see Heckman and Carneiro (2003) for a review).

choices and demographic factors. This provides a mapping from a set of initial conditions (that is, initial agents' distribution over states such as permanent and persistent idiosyncratic shocks and assets) into distributions over educational and economic attainments: this mapping turns out to be ideal to study the economic implications of alternative policy interventions.

3.2.1 Overview

We consider an economy where a unique good is produced, and it can be either consumed or used as physical capital. We specify an overlapping generations general equilibrium model for this economy that focuses on education and labour supply. Consumers maximise an intertemporal utility function over their finite life-cycle, with respect to education, labour supply and consumption/savings. Agents can accumulate assets representing ownership shares on physical capital. They have a maximum lifetime and they plan to consume their entire assets. However they may die before that leaving accidental bequests. The maximum possible lifetime is 99 years. Individuals can work up to age 65, and after that age they retire. They can also decide not to work before age 65 (by not supplying any labour). Retirement is financed by the accumulated assets. The population consists of 99 overlapping generations, each with an ex-ante identical distribution of heterogeneity.

Young and old households are not linked in any direct way. Bequests are pooled together and redistributed to all newly born individuals according to the steady state equilibrium wealth distribution. This reflects both inter-vivos transfers for education and actual bequests. Since we assume that assets must always be non-zero (liquidity constraints), these transfers are the only source of funding for education, other than possible government transfers. Education can only take place at the beginning of the life-cycle and the individual can attain one of three education levels, corresponding to less than high school, high school, and college. The costs of education consist of the opportunity cost, tuition fees net of any government subsidy and the psychic/utility costs of education. In addition individuals are endowed with different abilities which lead to different efficiency units of human capital and thus earnings. Thus wage differences among people are the consequence of differences in education (between group inequality) and differences in labor efficiency (within group inequality). People are perfect substitutes within schooling groups, regardless of their individual efficiency.

There is no aggregate uncertainty in the model. Once out of school the individual has to decide on the proportion of his/her time to devote to work and on consumption. All these decisions take place in an incomplete markets environment. The individual cannot borrow against human capital and faces uninsurable idiosyncratic wage shocks.

There is an aggregate production function with four inputs: the three levels of human capital, measured in total efficiency units supplied and physical capital. We solve the model as a closed economy with the interest rate determined endogenously.

The model is partly estimated and partly calibrated. First we estimate a wage process and extract relative prices for our three human capital measures. This allows us to compute the total supply of efficiency units of human capital in each of the three groups. From the estimation of the wage process we also estimate the stochastic process of wages, net of measurement errors which we take to be the process of uncertainty facing the individual.[15] The stochastic process of wages is taken to be education specific.

Next we estimate the aggregate production function which is taken initially to be Cobb-Douglas. However, we also carry out sensitivity analysis based on a number of different production function structures. Moreover, since there seems to have been a change in the production structure mainly due to skill biased technological change we use as our basis for the Cobb-Douglas specification the average shares over the period.

To obtain the parameters characterising preferences we use risk aversion coefficients taken from the literature and then we set the preference for labour supply to match the proportion of people working in the economy. Given these parameters we then calibrate the utility costs of education to match the proportions in each education group during 1978-82. The individual discount rate is calibrated to match the ratio of physical capital over total output.

[15]This may overestimate the degree of uncertainty; see Cunha, Heckman, and Navarro (2005).

3.2.2 Individual preferences

We use the index j to denote age. Agents have a probability to survive in each period, denoted as s_j, which is decreasing in age. Annuity markets are absent we use a random bequest mechanism to redistribute left over assets based on the prevailing equilibrium assets' distribution.[16]

Agents face educational choices based on returns and costs, which depend on age, asset holdings, permanent characteristics and labor shocks. Over the life cycle they choose the labor supply and consumption path maximizing expected lifetime utility.[17]

The period utility $u(c,l)$ is concave in consumption c and leisure $l = (1-n)$; it satisfies standard regularity conditions and in particular the Inada conditions. The education level is denoted by e, it takes three values, with $e = e_1$ the lowest and $e = e_3$ the highest. Permanent (unobserved) individual characteristics are denoted by θ and distributed over the domain $[\theta_{\min}, \theta_{\max}]$. We also assume that the distribution of ability θ is independent of time. We denote by $\{z\}_{j=1}^{\bar{j}}$ the sequence of uninsurable idiosyncratic shocks. Their law of motion is summarized by a stationary transition function π denoted as $\pi_{z_{j+1}|z_j} = \pi\{z_{j+1} \mid z_j\}$.

While in school individual utility is given by $u(c_j, f(\theta))$, where the function $f(\theta)$ reflects the psychic costs of schooling which may be thought of as leisure costs but may include other aspects of effort and like or dislike of the education process - hence we do not bound f to lie in the unit interval. These costs will depend on ability with the idea that more able individuals will suffer lower costs.

Given some initial values $\bar{x}_1$ for the state variables, household/individual utility over sequences of consumption and leisure, $c = \left\{c_1, ..., c_{\bar{j}}\right\}$ and $l = \left\{l_1, ..., l_{\bar{j}}\right\}$, as of age 1 is denoted $U(\bar{x}_1, c, l)$ and can be written as the expected discounted sum of

[16] Negative borrowing limits open up the possibility that people dying prematurely can be in debt. Yaari (1965) considers this case explicitly and proves that, with functioning credit markets not making systematic losses, the budget constraint must be such that individual can never go short on assets. We prefer to ignore this issue and let interest rates adjust appropriately.

[17] See discussion in chapter 2 for technical details regarding the agent's problem.

period utilities

$$U(\bar{x}_1, c, l) = E_{z\in Z}\left\{\sum_{j=1}^{j_{edu}} S_j\beta^{j-1}\left[\mathrm{d}_j u(c_j, f(\theta)) + (1-\mathrm{d}_j)u(c_j, l_j)\right] + \sum_{j=j_{edu}}^{\bar{j}} S_j\beta^{j-1}u(c_j, l_j)\right\} \tag{3.1}$$

where d_j is a binary variable which is 1 if the agent is in education and 0 otherwise, j_{edu} denotes the last age of education, $S_j = \left(\prod_{i=1}^{j} s_i\right)$ denotes the probability of surviving to age j and β is the intertemporal discount factor. For the first j_{edu} ages the individual may decide to be in education which is why there are two alternative forms for the utility function depending on the action taken. We restrict all education choices to take place in the beginning of life.[18] Note that once in education $f(\theta)$ is fixed and only depends on ability θ. The period budget constraint is given by

$$\begin{aligned} c_j + a_{j+1} &= \left[1 + r(1-\tau_k)\right]a_j + w_e e^{\epsilon_j} n_j(1-\tau_{n^e})(1-\mathrm{d}_j) + \\ &\quad - (D_e - Tr_e)\,\mathrm{d}_j \end{aligned} \tag{3.2}$$

where a_j denotes individual asset holdings[19] at age j and r is the risk-free interest rate. For the purposes of policy analysis we distinguish between the taxation of capital income τ_k and the taxation of labour income τ_{n^e}.[20]

The D_e is the direct cost of schooling and Tr_e summarizes government subsidies towards education e. The term e^{ϵ_j} denotes individual labour efficiency, with ϵ_j defined as

$$\epsilon_j(\theta, e, z) = \theta + \xi_j(e) + z_j \tag{3.3}$$

where $\xi_j(e)$ is an education-specific age profile.

[18] This restriction reduces the computational burden significantly

[19] Individual asset holdings satisfy: $a_j \geq a_{\min}$ for every j and $a_{\bar{j}+1} \geq 0$. The first inequality is a borrowing constraint, whereas the second is a transversality condition for agents reaching age $\bar{j}$.

[20] Heckman (1976) first noted the importance of this distinction when considering investments in Human Capital. Changing the cost of intertemporal substitution will affect investment decisions.

3.2.3 Solving the individual problem

The individual's problem may be solved recursively by backwards induction. Denote by x_j the value of the state variables at age j and by $W_j(x_j)$ the optimum value function at age j. The state vector includes the current value of the shock z, which is assumed to arrive at the beginning of the period, as well as permanent characteristics, past values that are relevant for predicting future outcomes and, of course, current assets and education levels. Following the age of life when no more education choices can be made the individual chooses consumption and labour supply to solve the simpler problem

$$W_j(x_j) = \max_{\{c_j,n_j\}} u\left(c_j, 1-n_j\right) + s_j\beta \int_Z \pi_{z_{j+1}|z_j} W_{j+1}\left(x_{j+1}\right) dz_{j+1} \tag{3.4}$$

subject to the budget constraint (3.2) with $\mathrm{d}_j = 0$ and subject to the constraint $a_{j+1} \geq a_{min}$ where a_{min} is some exogenous minimum level of assets.[21]

Previously, during the early ages of life, the individual's problem is complicated by the fact that she/he needs to decide on whether to obtain education. When education is still an option[22] the problem solved is

$$\begin{aligned} W_j(x_j) = & \max_{\{c_j,n_j,d_j\}} \left\{\mathrm{d}_j u\left(c_j, f(\theta)\right) + (1-\mathrm{d}_j) u\left(c_j, l_j\right)\right\} + \\ & + s_j\beta \int_Z \pi_{z_{j+1}|z_j} W_{j+1}\left(x_{j+1}\right) dz_{j+1} \end{aligned} \tag{3.5}$$

subject to the budget constraint, the asset constraint mentioned above and subject to the constraint that $n = 0$ if $\mathrm{d}_j = 1$, i.e. we do not allow for work and education at the same time.

3.2.4 Aggregate variables

We study equilibrium allocations and assume a stationary population. The aggregate states of the economy are physical capital K and efficiency-weighted aggregate labour

[21] We set $a_{min} = 0$ in the numerical experiments

[22] See chapter 2 for a description of the timing of education decisions in this model.

supplies (referred to as human capital aggregates) H_1, H_2, and H_3.

We define the measure space $(X, \mathcal{F}(X), \psi_j)$, where X is the individual state space.[23]

For each set $Q \subseteq \mathcal{F}(X)$, let ψ_j represent the normalised measure of age j agents whose individual states lie in Q as a proportion of all age j agents. Calling ζ_j the fraction of age j agents in the economy we define

$$\mu = \mu(Q, j) = \zeta_j \psi_j(Q)$$

which is a measure of agents belonging to age group j with individual state vector $(\theta, e, z, a) \in Q$.[24]

The aggregate states determine the relative prices in the economy. In steady state there will be no change in the relative wages between the three different types of human capital. However, the feature of the model that allows for such relative price variation is key because as the policy alters the supply of each education group, relative prices will change and this will lead to different steady state levels of supply for H_1, H_2, and H_3. The total stock of human capital of type e is the sum of the efficiency weighted individual labor supplies of type e which are defined by

$$h_j(\theta, e, z, a) = e^{\epsilon_j(\theta, e, z)} n_j(\theta, e, z, a) \tag{3.6}$$

The total stock of human capital of type e is denoted as H_e is defined as

$$H_e = \sum_j \zeta_j \int_X h_j(x)\, d\psi_j(x) = \sum_j \zeta_j \int_X \epsilon_j(\theta, e, z)\, n_j(x)\, d\psi_j(x)$$

where $\psi_j(x) = \psi_j(\theta, e, z, a)$.

The demographics are stable, so that age j agents make up a constant fraction ζ_j of the population at any point in time. The ζ_j values are normalized to sum up to 1

[23] $X = \Theta \times \Im \times Z \times \bar{A}$ and $\mathcal{F}(X) = \mathcal{F}(\Theta) \times \mathcal{F}(\Im) \times \mathcal{F}(Z) \times \mathcal{F}(\bar{A})$ is its sigma-algebra.

[24] μ is a measure on $(\Gamma, \mathcal{F}(\Delta))$, where $\mathcal{F}(\Delta)$ is the Borel σ-algebra on $\Delta = \Upsilon \times X = \Upsilon \times \Theta \times \Im \times Z \times \bar{A}$, defined as $\mathcal{F}(\Delta) = \mathcal{F}(\Delta) \times \mathcal{F}(\Theta) \times \mathcal{F}(\Im) \times \mathcal{F}(Z) \times \mathcal{F}(\bar{A})$. Ergo, for any given $Q \in \mathcal{F}(\Delta)$, $\mu(Q)$ indicates the mass of agents whose individual state vectors lie in Q.

and are such that $\zeta_{j+1} = s_j \zeta_j$.

3.2.5 Market structure

Our setup is an incomplete markets one where idiosyncratic risk cannot be insured, other than by self-insurance through precautionary savings. However we also impose liquidity constraints, which will be biting for the more able young people, unless they have inherited wealth. There is no aggregate uncertainty. The unique physical good is used as numeraire. All unemployment in the model is voluntary.

With missing annuity markets the assets left behind by agents who die at age j are distributed to the youngest age group according to the density law prevailing among age j agents.[25]

Given differential mortality and life-cycle assets savings the cross-sectional distribution of bequests changes with age. The newborns' asset density is assumed to be equal to the aggregate distribution of bequests[26]

$$\psi_1(a) = \sum_{j=1}^{\bar{j}-1} \frac{\zeta_j (1 - s_j)}{\sum_{j=1}^{\bar{j}-1} \zeta_j (1 - s_j)} \psi_{j+1}(a)\, da \tag{3.7}$$

where $\psi_j(a)$ denotes the age j marginal assets density[27]

Let q_j denote individual bequest at age j. The bequest mechanism described above is such that[28]

[25] This bequest mechanism has the desirable feature of making the age 1 assets density depend on the older ages assets densities generated in equilibrium.

[26] The only restriction imposed on the distribution of assets at age 1 is that no agent should be born in debt. This is achieved by properly rescaling the lower tail and the average of the bequests' distribution

[27] This is defined as

$$\psi_{j+1}(a) = \int_X \psi_{j+1}(\theta, e, z, a)\; d\theta dedz$$

[28] In this chapter we will provide results for the case in which newborns have independent draws in ability and assets. However, if one knew the correlation between permanent characteristics and initial wealth, it would be possible to introduce dependence between the ability and asset draws.

$$E(q_1) = \int_{\bar{A}} \psi_1(a)\, a_1 da \tag{3.8}$$

$$q_j = 0 \quad \text{for} \quad j = 2, ..., \bar{j}$$

and the amount of wealth that is bequitted in each period is[29]

$$\zeta_1 \int_{\bar{A}} \psi_1(a)\, a_1 da = \sum_{j=1}^{\bar{j}-1} \zeta_j (1 - s_j) \int_{\bar{A}} a_{j+1} \psi_{j+1}(a)\, da$$

3.2.6 Technology

Firms maximize profits using a CRS technology and set wages competitively. The aggregate technology employs physical and human capital and is denoted as $F(H, K)$ with $H = \{H_1, H_2, H_3\}$. The relationship between human capital factors (H) and physical capital is expressed as a Cobb-Douglas :

$$F(H, K) = \bar{A} H^{1-\alpha} K^{\alpha} \tag{3.9}$$

$\bar{A}$ is a TFP coefficient and the isoelastic, general definition of the HC input is

$$H = \{A_1 H_1^{\rho} + A_2 H_2^{\rho} + A_3 H_3^{\rho}\}^{\frac{h}{\rho}} \tag{3.10}$$

with $h = 1$ given the CRS assumption.[30] This specification allows for the elasticity of substitution to differ between unskilled labour H_1 and the other two inputs.

In this specification (A_1, A_2, A_3) are share parameters, while ρ pins down the Allen elasticity of substitution among different weigthed labour inputs. In the CES case the Allen elasticity of substitution between any two inputs is $\frac{1}{1-\rho}$. [31] When ρ is equal to

[29] In reality, only a part of intra-family wealth transfers are intra-vivos. For a discussion of related issues see Gale and Scholz (1994).

[30] For strict quasi-concavity of the production function ρ has to lie within $(-\infty, 1)$.

[31] The Allen partial elasticity of substitution is also known as the Allen/Uzawa elasticity and is the most widely used. However, Blackorby and Russell (1989) show that there is no intution about what this measures. Blackorby and Russell advocate the use of the so-called Morishima elasticity. Another alternative for multisector models would be the so-called direct elasticity of substitution proposed by McFadden (1963). In what follows we just use the Allen elasticity as a simple approximation.

zero the technology is Cobb-Douglas, whereas values of ρ greater than zero indicate more substitutability than in the Cobb-Douglas case.

An alternative specification could be[32]

$$H = \left\{ A_1 H_1^{\rho_1} + [A_2 H_2^{\rho_2} + A_3 H_3^{\rho_2}]^{\frac{\rho_1}{\rho_2}} \right\}^{\frac{1}{\rho_1}}$$

which allows for the elasticity of substitution to differ between unskilled labour H_1 and the other two inputs.[33] This specification also has a symmetry property imposing that the elasticity of substitution between H_2 and H_3 is the same as the that between H_3 and H_1. Therefore, if $\rho_2 > \rho_1$ we have that H_3 is more complementary with H_1 than with H_2. Also, the grouping allows separate parts of the above technology to be Cobb-Douglas, when either ρ_2 or ρ_1 tend to zero.

In practice we have not been able to obtain meaningful estimates of the more general, unconstrained production functions and we present estimates for a Cobb-Douglas specification with $\rho_1 = \rho_2 = 0$. However we do use the isoelastic general specification as a basis for sensitivity analysis in the simulations. In chapter 4 we present an alternative method which seems able to estimate one type of unrestricted technology with more precision.

The equilibrium conditions require that marginal products equal pre-tax prices so that $w_e = \frac{\partial F}{\partial H_e}$ for any education level e, and $r + \delta = \frac{\partial F}{\partial K}$, where δ is the depreciation rate for capital.

3.2.7 Government

Government has revenues from proportional taxation of labor and asset income at respectively τ_{n^e} and τ_k rate, and uses part of these revenues to subsidize education via a transfer Tr_e. We call G the residual non-education general government expenditure and assume that G is lost in non-productive activities. The government's behaviour is fully described by the budget constraint, which requires that expenditures equal

[32]This specification has been attributed to Sato (1967) by Hamermesh (1993).

[33]See Caselli and Coleman (2006) for a discussion.

revenues obtained from taxation[34]. The government has a balanced budget in each period.

3.3 Equilibrium

We use a notion of equilibrium in which the state variables' distribution remains unchanged over time. This notion of equilibrium is known as stationary recursive competitive equilibrium, see Lucas (1980). In the Appendix to chapter 3 there is a brief description of the steps required to define a stationary measure ψ_j, such that $\mu(x,j) = \zeta_j \psi_j(x)$ is stationary, as a function of the markov process $\pi\{z_{j+1} \mid z_j\}$ and of the decision rules $\mathrm{d}_j(x)$ and $a_{j+1}(x)$, where $x \in X$ is a vector of state variables.

Let $(X, \mathcal{F}(X), \psi_j)$ be an age-specific measure space with state space X and $\mathcal{F}(X)$ be a σ-algebra on X.

Given some state vector $x \in X$, a stationary recursive competitive equilibrium for this economy is a set of decision rules $\mathrm{d}_j(x)$, $a_{j+1}(x)$, $c_j(x)$, $n_j(x)$, value functions $W_j(x)$ and $V_j(x)$, price functions w_e, r, densities $\left(\psi_1, ..., \psi_{\bar{j}}\right)$ and $\left(\zeta_1, ..., \zeta_{\bar{j}}\right)$, and a law of motion Q, such that:

1. $\mathrm{d}_j(x)$, $a_{j+1}(x)$, $c_j(x)$ and $n_j(x)$ are optimal decision rules and solve the household's problem;
2. $W_j(\theta, e, z, a)$ and $V_j(x)$ are the associated value functions;
3. Firms employ inputs so that

$$w_e = F_{H_e} \quad for \;\; e \in \Im$$
$$r + \delta = F_K \;;$$

[34]The government budget constraint is

$$G + \sum_j^{\bar{j}} \zeta_j \int_X Tr_e \mathrm{d}_j(x)\, d\psi_j(x) =$$
$$= \sum_j \zeta_j \int_X [1 - \mathrm{d}_j(x)]\, \tau_{n^e} w_e h_j(x)\, d\psi_j(x) + \sum_j \zeta_j \int_{\bar{A}} r\tau_k a_j\, d\psi_j(a)$$

We assume that the government has a balanced budget in each period.

4. $\psi_j(x)$ is a stationary measure, that is $\psi_j(F) = Q(F, \psi_j)$, where $Q(\cdot,\cdot)$ is the law of motion of $\psi_j(\cdot)$ and is generated by the optimal decisions $\mathrm{d}_j(x)$, $a_{j+1}(x)$, $c_j(x)$;[35]

5. The good, asset and labour markets clear.[36]

The goods market clearing equation is derived by integrating the individual budget constraint.

3.4 Estimation

With the exception of the parameters for intertemporal substitution the remaining ones are obtained by a combination of estimation and calibration using data from the US.

We thus estimate the wage process, the distribution of unobserved heterogeneity and the aggregate production function. We then calibrate the costs of education function $f(\theta)$ to fit the distribution of education by ability in the four year period from 1978 to 1982. Our sources of data are the CPS, the PSID and NIPA.

3.4.1 Estimating wage equations: skill prices 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 extract from the data the distribution of unobserved heterogeneity affecting wages and education choices as well as the stochastic process of the shocks.

We start by specifying an education specific wage equation for individual i wages in period j, w_{eit}

$$\ln w_{eit} = w_{et} + g_e(age_{eit}) + u_{eit} \tag{3.11}$$

[35] Given ζ_j, also $\mu(x,j) = \zeta_j \psi_j(x)$ is a stationary measure.

[36] Equilibrium definitions in the asset and good markets must include cross border asset holding FX if the interest rate r is constant. The Appendix to chapter 3 contains a derivation of the market clearing equations.

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

$$u_{eit} = \theta_i + z_{eit} + m_{it} \tag{3.12}$$

where θ_i represents unobserved fixed effects, z_{eit} is the (persistent) shock to wages and m_{it} is measurement error, assumed *iid.* Self-selection implies that fixed effects are correlated with both education decisions and observed wage rates. However, a within groups transformation eliminates the source of self-selection and identifies the changes in the returns to education over time as well as the way wages grow with age by education group. Thus we estimate by OLS

$$(\ln w_{eit} - \ln \overline{w}_{ei}) = (\ln w_{et} - \ln \overline{w_e}) + g\left(age_{eit} - \overline{age}_{ei}\right) + (u_{eit} - \overline{u}_{ei}) \tag{3.13}$$

where the upper-bar denotes an (individual) time average and where g is a polynomial of order two for the lowest education group and of order 4 for the two higher education groups. The term $(\ln w_{et} - \ln \overline{w_e})$ is modelled as time dummies. The residuals from this equation can be used to identify the persistence of wage shocks, and we discuss this below.

3.4.2 Wage data and results

For the estimation of wage equations we use longitudinal data from the PSID. The sample is based on annual interviews between 1968 and 1997 and on bi-annual interviews from 1999 onwards. 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.

The main earnings' variable in the PSID refers to the head of the household[37] and is

[37] 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

described as total labor income of the head.[38] 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.

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). A detailed description of our sample selection is reported in the appendix: in brief, we select heads of household aged 25-60 who are not self-employed and have positive labor income for at least 8 (possibly non continuous) years.

The age polynomials from the wage equation are presented in table (3.1).

Table 3.1: Age polynomials

Dependent variable: log hourly earnings		
coeff.	point estimate	S.E.
	Education=LTHS	
age	0.0412505	0.0081143
age^2	-0.0004179	0.0000905
	Education=HSG	
age	0.4928285	0.1071015
age^2	-0.0162768	0.0039883
age^3	0.0002413	0.0000644
age^4	-0.00000134	0.00000038
	Education=CG	
age	0.8697329	0.1560285
age^2	-0.0282	0.0058548
age^3	0.0004149	0.0000953
age^4	-0.0000023	0.00000057

Figure (4.2) plots the age profiles implied by the polynomial estimates for different education groups.

By fitting the within group specification of the wage (log of hourly earnings) equation we also obtain $\ln(\hat{w}_{et})$, estimates of the growth of log-price of labor by education

estimates.

[38]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.

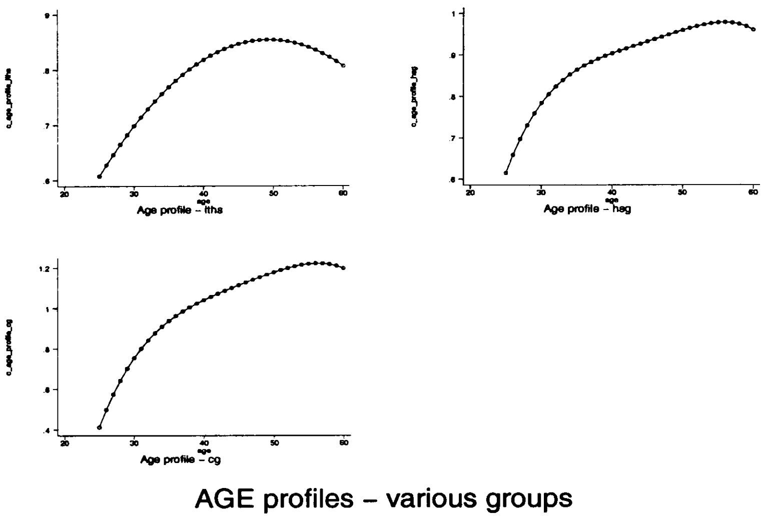

Figure 3.1: Age profiles of labor efficiency by education group

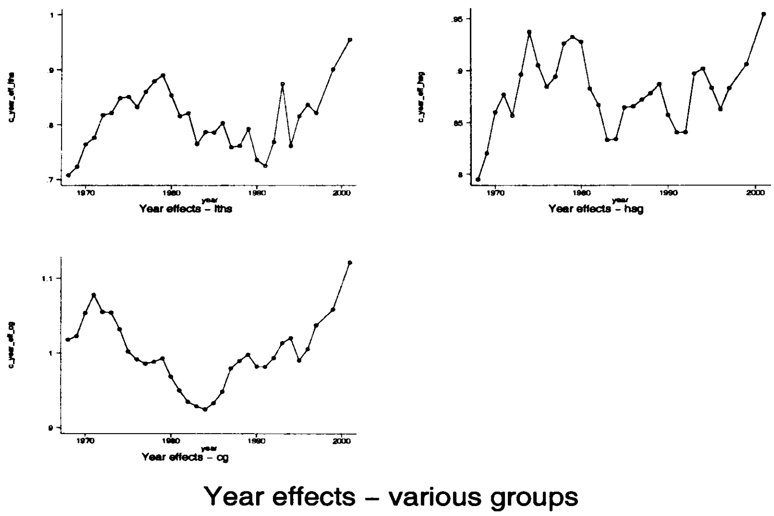

Figure 3.2: Log of marginal labor productivity, by education group

and year, which are plotted in figure (3.2). The time effects have a natural interpretation as time varying prices of skills associated to different education groups. The fact that the relative prices vary this much is a key justification for treating the different education levels as different types of human capital.

3.4.3 Permanent characteristics and their distribution

For the purposes of simulation we require the unconditional distribution of permanent characteristics (ability) as reflected by the fixed effect θ_i. We thus use the estimate

$$\widehat{\theta}_i = \frac{\sum_{j=1}^{J(i)} \ln w_{it} - \widehat{\ln w_t} - g\left(\widehat{age_{it}}\right)}{T(i)}$$

where $T(i)$ is the total number of observation available on agent i. If we assume that the unconditional distribution of ability has not changed over the time period covered by our sample, we can use the estimated fixed-effects as an estimate of the $\{\theta_i\}$ distribution over the working population.

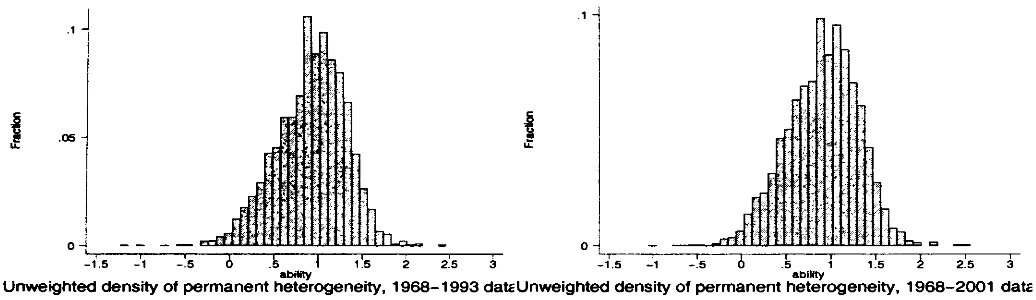

Figure 3.3: Log density of fixed effects for 1967-1993 and for 1967-2000

In figure (3.3) we report the empirical frequencies of $\widehat{\theta}$ obtained by aggregating both cross-sectionally and longitudinally.

The estimation variance of $\widehat{\theta}_i$ will inflate the overall variance of unobserved heterogeneity. To mitigate this problem we have traded off some representativeness by taking individuals who are observed for at least eight periods.

3.4.4 Analysis of labour efficiency shocks

We now use the residuals from the wage equation to estimate our assumed stochastic process for wages. First note that we can treat as observable the following:

$$u_{eit} = \ln w_{eit} - g_e\left(age_{eit}\right) - \ln w_{et} - \theta_i \tag{3.14}$$

We assume that u_{eit} can be decomposed into two components

$$u_{eit} = z_{eit} + m_{eit}$$

where z_{eit} is an autocorrelated error process and m_{eit} is classical measurement error, $iid\left(0, \sigma_{em}^2\right)$, and where $\{z_{ei}\}_t$ is a autocorrelated process with education specific parameters

$$z_{eit} = \rho_e z_{eit-1} + \varepsilon_{eit}$$

in which $\varepsilon_{eit} \sim iid\left(0, \sigma_\varepsilon^2\right)$. We can achieve identification of the autoregressive parameters in one of several ways. With an external estimate of the measurement error variance we can use the following expressions to estimate σ_ε^2 and ρ :

$$\rho = \frac{COV\left(z_{eit}, z_{eit-1}\right)}{VAR\left(z_{eit}\right)} = \frac{COV\left(u_{eit}, u_{eit-1}\right)}{VAR\left(u_{eit}\right) - VAR\left(m_{it}\right)} \tag{3.15}$$

$$VAR\left(u_{eit}\right) = \frac{\sigma_{\varepsilon_e}^2}{1 - \rho_e^2} + \sigma_m^2$$

where we can substitute the covariances of u with sample analogues. However it is also possible to use the variance of u and its first two auto-covariances to identify the variance of the measurement error as well. Thus we have that

$$\rho_e = \frac{COV\left(u_{eit}, u_{eit-2}\right)}{COV\left(u_{eit}, u_{eit-1}\right)}$$

and the rest follows immediately. In practice we replace the error terms u with the residuals for the wage regression as defined in (3.14).

Results for the wage process

We present estimates of the autoregressive coefficients obtained using external estimates of measurement error by French (2000), who provides a lower and a upper bound estimate for measurement error (respectively 0.0172 and 0.0323). Our results are based on an average of the two. The (bootstrapped) standard errors are in parentheses.

Table 3.2: Autoregressive coefficient $\hat{\rho}$. Bootstrapped S.E. in parenthesis

Group 1	Group 2	Group 3	Pooled
0.651	0.557	0.608	0.584
(0.130)	(0.042)	(0.058)	(0.034)

These findings are apparently in contrast with some of the recent literature, among others Storesletten, Telmer, and Yaron (2004) and Meghir and Pistaferri (2004). However, using the upper estimate of measurement error we get parameters much closer to one. Of course, with near unit-root persistence of wage shocks the identification of fixed effects would suffer from severe initial conditions problems (for a discussion of incidental parameters problem see Heckman (1981)). In chapter 4 we take a more orthodox stance and use an exogenous distribution of ability based on 1972 PSID test scores while assuming unit-root behaviour of labor efficiency shocks.

Table(3.3) presents the point estimates of σ_{ε}^2.

Table 3.3: Variance of Ar(1) perturbation (σ_{ϵ}^2). S.E. in parenthesis

Group 1	Group 2	Group 3	Pooled
0.024	0.054	0.099	0.063
(0.011)	(0.005)	(0.012)	(0.006)

3.4.5 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 it also requires some measure 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.[39] 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 and we can normalise one of the initial prices to one. Now note that we have one degree of freedom. We can set the initial relative price of high school and of college graduate labour and we can then choose the 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.[40] 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.[41]

[39]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).

[40]Eliminating all zero-earnings observations rules out the possibility to incorporate employment risk, which is possibly an important source of risk.

[41]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

Figure (3.4) reports the number of people working in each year by education group, as reported by the CPS.

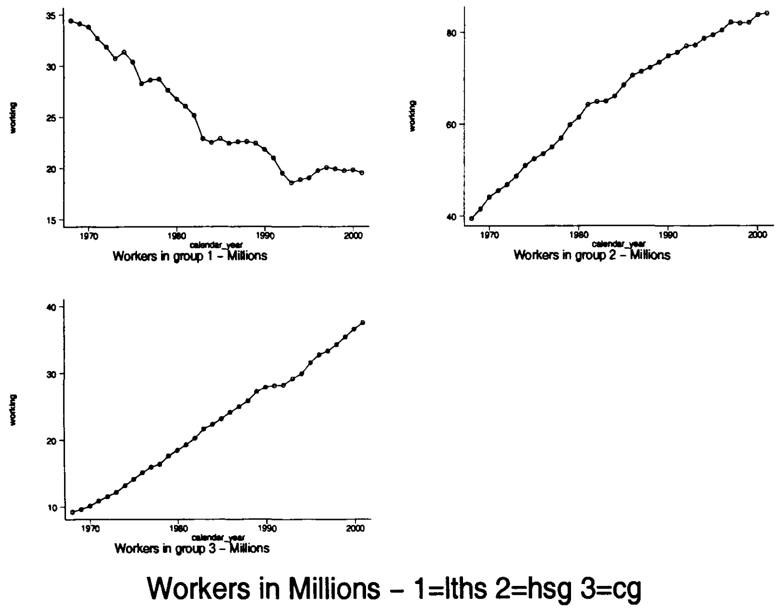

Figure 3.4: Employed workers in millions, by education

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

Figure (3.5) plots both the total wage bills in billions of dollars whereas figure (3.6) plots averages. 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 (West (1985)) which is based on the original Hill 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).

undisclosed and confidential non top-coded data available only at the BLS.

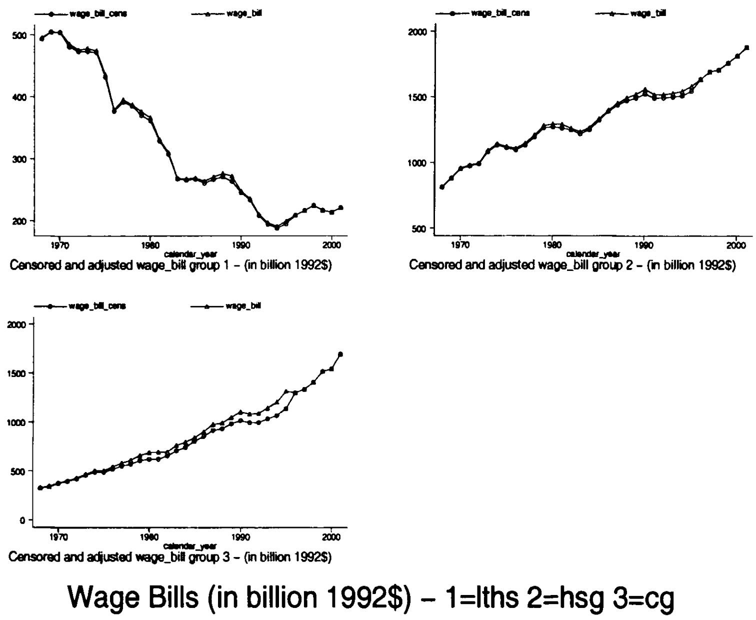

Figure 3.5: Total earned labour income, by education, in billions of 1992 dollars.

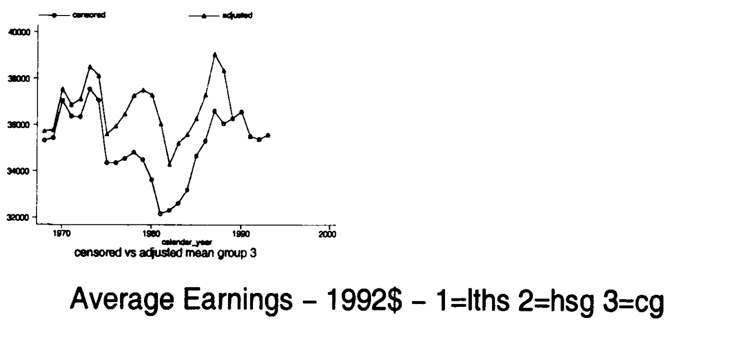

Figure 3.6: Average earned labour income, by education, in 1992 dollars.

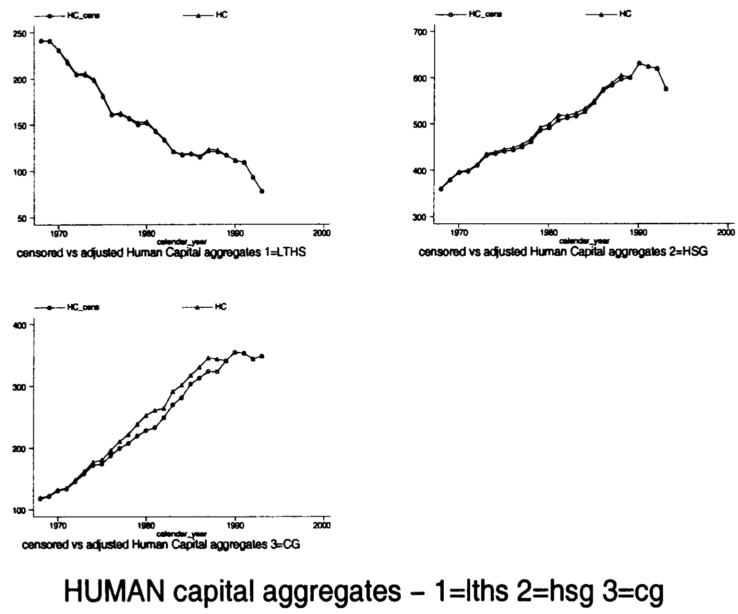

Figure 3.7: Efficiency weighted labor supply (HC) in billions of 1992 dollars, by education

Finally, dividing the wage bills by the exponentiated value of the time effects estimated through the wage equations using PSID data we finally obtain point estimates of the value of efficiency weighted total labor supply (human capital aggregates) by education and year. These are plotted in figure (3.7).

Notice that the evolution of human capital over time is non-monotonic, unlike the wage bills for the two highest education groups. This is due to the large increase in the level of estimated marginal product of these two factors in the early 1990s, which has grown proportionally more than the total remuneration of these factors.

3.4.6 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

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

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.

Share parameters A and B can be easily expressed as a function of the aggregate wage bills. We can then compute the output shares of College labour, High School labour and Dropouts labour, which are respectively A, $(1-A)B$ and $(1-A)(1-B))$ in terms of technology parameters.

Applying this procedure separately for each year we can pinpoint the evolution of these functions over the sample period.

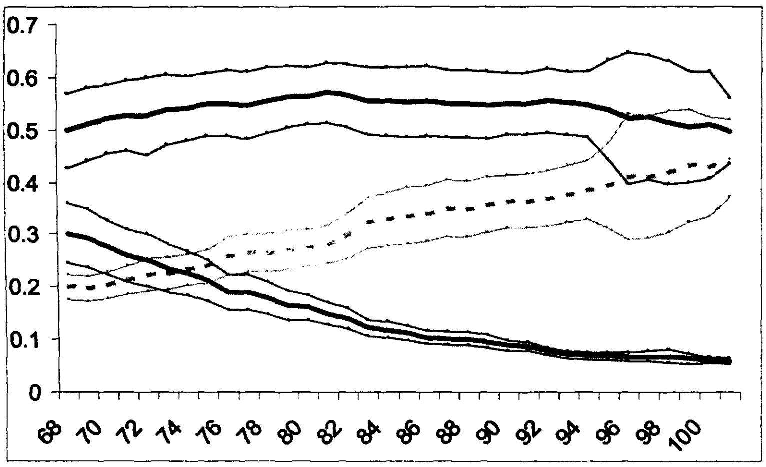

Figure 3.8: Labor shares in human capital input of technology, computed using Cobb-Douglas specification (with bounds equal to +/- 2 standard errors). Period: 1968-2000. Larger bounds after 1996 are due to changes in top-coding of income in the CPS.

Figure (3.8) reports the value of the share parameters (with bounds equal to 2 standard errors) associated to each human capital variety. In figure (3.8) the upward sloping line represents the College output share, whereas the downward sloping one represents the Dropouts output share. The flat line on top refers to the output share for High School graduates.

The time average of such shares is $A = 0.33$, $(1-A)\,B = \ 0.54$ and $(1-A)\,(1-B) = 0.14$. 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 the strong changes in the education composition of the labour force, confirm what we already noticed in terms of marginal productivity of labour using PSID data: major shifts in technology have taken place between the late 1960s and the end of the century.

We follow up our initial findings by performing some additional inference on the technology parameters. In order to do this we first approximate the total human capital factor $H = F\{H_1, H_2, H_3\}$ by combining NIPA and CPS data on wage bills and physical capital[42] and then use a 2-step GMM method which controls for endogeneity and serial correlation of TFP to estimate the parameters. We use lagged shares as instruments. .

The results of the GMM estimation of our favourite specification for the log-linearized C-D technology are reported in the following tables (standard errors in parenthesis) for two alternative moment weighting matrix choices (the identity matrix and the optimal matrix). In the Appendix to chapter 3 we report results for all other specification.

Table 3.4: Long-term (1967-1997) labour output shares. S.E. in parenthesis

	First Step Weighting: Identity Matrix	First Step Weighting: Optimal Matrix
A	0.260	0.284
	(0.200)	(0.207)
B	0.783	0.790
	(0.115)	(0.123)

We also find that the linear trend included to control for TFP deterministic variation is estimated to be 0.035 and strongly significant. Given the log specification, this is equivalent to an average annual TFP growth rate of roughly 3.5% between 1967 and 1997, a value that lies on the higher end of current estimates for the period.

The point estimates for A and B give labor shares very similar to the long-term averages we estimate using the initial Cobb-Douglas computation. The labor shares roughly sum up to one, even though we do not impose this restriction in the estimation procedure.

[42]Details of the model and estimation procedure can be found in the Appendix to chapter 3.

3.5 Simulations

Estimation above has yielded the stochastic process for wages, the distribution of ability, the life-cycle growth of wages and finally the aggregate production function. In the next step we set the preference parameters and we choose the utility cost of education to match as close as possible the proportions completing each of the three levels of education.

3.5.1 Preferences parameters

Wealth and consumption data are probably not of sufficient quality to estimate a joint model of labour supply, consumption and education choice, particularly because we will have to observe all forms of wealth reliably, together with consumption. Thus we decided to rely on earlier Euler equation estimates for the preference parameters, as well as matching aggregate labour supply levels. Thus we specify the utility function to be of the CRRA type, i.e.

$$u\left(c_j, l_j \mid \mathrm{d}_j = 0\right) = \frac{\left[c_j^{\nu} l_j^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda}$$

$$u\left(c_j \mid \mathrm{d}_j = 1\right) = \frac{\left[c_j^{\nu} f^{e}(\theta)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda}$$

The parameters ν and λ of the period utility jointly pin down the intertemporal elasticity of substitution of consumption $\frac{1}{1-\nu(1-\lambda)}$ (ISE) as well as the level of labour supply over the lifecycle. Se set the ISE to 0.75 as in Blundell, Browning, and Meghir (1994) and Attanasio and Weber (1993). Given this, a value of $\nu = 0.33$ and hence $\lambda = 2.00$ matches the labour supply data very well.

3.5.2 Demographic and cost parameters

Individuals are assumed to be born at the real age of 16, and they can live a maximum of $\overline{j} = 84$ years, after which, at the real age of 99, death is certain (retirement occurs after age 60). The sequence of conditional survival probabilities $\{s\}_{j=1}^{99}$ is based on mortality tables for the US and we do not differentiate mortality rates by sex or race.

The direct cost of education D_e is set to be equal to 0.3 times the average earnings in the economy, which corresponds to an estimate of average (in-state) tuition costs for public and private colleges in the US.[43]

Tuition subsidies (Tr_e) as a share of average earnings have changed over the last 30 years. A long term average stands at roughly 1/2 of the tuition costs.

3.5.3 Some simple tuition experiments

The numerical experiments we report in the rest of this section are compared to a simple benchmark economy in which the discount factor β is set to match a physical capital over output ratio of 3.0. The resulting discount factor is very close to one, with the first 3 decimals all being equal to 9. The depreciation rate is set to 0.07, which we compute from NIPA data. No negative assets are allowed in the benchmark economy. The deterministic leisure function $f^e(\theta)$ is discretised and then calibrated to approximately match enrolment rates within different ability bins.[44]

The initial wealth distribution is endogenously determined in this simulations: the accidental bequests are distributed to the new-borns following the steady state asset density. Thus some people are born with zero assets and others with different, positive amounts, which implies that some will be facing a tight liquidity constraint for college education. In the simulations we are not correlating initial assets and permanent characteristics, although we plan to do it in the future.

The tuition subsidy experiment is implemented by giving people, ceteris paribus, a transfer (same for all) equal to a percentage of the direct cost of schooling. The extra costs are covered by extra proportional labour income taxation.

The top panel of table (3.5) shows the results for the benchmark economy. The bottom panel shows what happens in partial equilibrium, when prices for human capital do not adjust. However, taxes must change to fund the tuition subsidy and of course the underlying wealth distribution does change as well as the work behaviour. The middle panel shows the general equilibrium results where human capital returns and interest rate are allowed to change.

[43]Source: Education digest, NCES, National Center for Education Statistics.

[44]If we keep the utility costs fixed enrolment shares are sensitive to the choice of aggregate technology

In partial equilibrium this universal subsidy leads to a substantial increase in college graduates from 20% to 25%. When breaking this down by ability we see that the increase is high for all ability groups, relative to their original position. In addition, this seems to have come for almost "free" since the tax on labour only needs to increase marginally. This is because the policy attracts a number of high ability and previously liquidity constrained individuals into higher education.[45] They earn high levels of income which more than compensate the cost of educating them. In fact there is a substantial increase in the college level human capital aggregate from 5.41 to 6.5. This is precisely the logic underlying a number of educational subsidy programmes around the world. Thus in partial equilibrium, the policy pays for itself.

In General Equilibrium though the situation is quite different, at least as far as the aggregate shares are concerned. Following the policy there is a very small decline in aggregate college attendance. This is due to the decline in the marginal product of college level human capital. However, the aggregate figures hide important differences within ability groups. These show a decline in College attendance *vis a vis* the baseline for ability levels two and three and an increase in College attendance for the highest ability level 4. In addition there is an increase in the rates of high school graduation for the lowest level of ability in response to an increase in the relative price for high school graduates. Finally there is a decline in college for the second ability group. All this adds up to an increase in the supply of human capital for the lowest and highest education groups: the subsidy has in fact led to an increase in inequality.

Similar results have been obtained when we use a production function with a higher elasticity of substitution see table 3.6.

3.6 Conclusions

We combine estimation and calibration to obtain an overlapping generations General Equilibrium model with heterogeneous agents and idiosyncratic uncertainty. Individuals choose education levels, labour supply and consumption within an incomplete

[45] Admittedly, this is also due to the tight zero borrowing limit that has been imposed for all agents. In this setup people who are born with zero assets are kept out of education.

Table 3.5: Simulation results. Technology 1: Cobb-Douglas

GROUPS

Benchmark (Tuition $3105 = 30% of median income)

	Edu. Participation (aggr.shares)			Human Capital Aggregates		
	Less than HS	HS	College	Less than HS	HS	College
	0.34	0.46	0.20	3.99	9.30	5.41
	Edu. Shares by ability			Marg. Products after Tax. and Depr.		
Ability 1 (lowest)	0.94	0.055	0.004	0.56	1.0	0.95
Ability 2	0.36	0.53	0.11	Aver. Post-Tax Labour Earn.-1990 US$		
Ability 3	0.23	0.52	0.26	668	1753	2122
Ability 4 (highest)	0.19	0.43	0.38			
% with wealth=0	0.082	r	0.0252	Tax lab	0.27	

General Equilibrium (50% Subsidy)

	Edu. Participation (aggr.shares)			Human Capital Aggregates		
	Less than HS	HS	College	Less than HS	HS	College
	0.34	0.47	0.19	4.54	8.67	5.43
	Edu. Shares by ability			Marg. Products after Tax. and Depr.		
Ability 1 (lowest)	0.92	0.075	0.005	0.53	1.0	0.94
Ability 2	0.36	0.55	0.092	Aver. Post-Tax Labour Earn.-1990 US$		
Ability 3	0.24	0.52	0.24	652	1710	2155
Ability 4 (highest)	0.20	0.38	0.42			
% with wealth=0	0.082	r	0.02519	Tax lab	0.28	

Partial Equilibrium (50% Subsidy)

	Edu. Participation (aggr.shares)			Human Capital Aggregates		
	Less than HS	HS	College	Less than HS	HS	College
	0.34	0.42	0.25	4.29	8.01	6.50
	Edu. Shares by ability			Marg. Products after Tax. and Depr.		
Ability 1 (lowest)	0.94	0.050	0.011	0.5575	1.0	0.95
Ability 2	0.36	0.47	0.17	Aver. Post-Tax Labour Earn.-1990 US$		
Ability 3	0.23	0.47	0.30	668	1767	2075
Ability 4 (highest)	0.19	0.38	0.43			
% with wealth=0	0.086	r	0.0252	Tax lab	0.272	

Table 3.6: Simulation results. Technology 2: E.of.S.=1.54

GROUPS

Benchmark (Tuition \$2929 = 30% of median income)

	Edu. Participation (aggr.shares)			Human Capital Aggregates		
	Less than HS	HS	College	Less than HS	HS	College
	0.34	0.50	0.16	4.48	9.71	4.46
	Edu. Shares by ability			Marg. Products after Tax. and Depr.		
Ability 1 (lowest)	0.89	0.10	0.011	0.44	1.0	0.94
Ability 2	0.38	0.56	0.058	Aver. Post-Tax Labour Earn.-1990 US$		
Ability 3	0.24	0.55	0.21	563	1816	2310
Ability 4 (highest)	0.19	0.45	0.36			
% with wealth=0	0.082	r	0.025	Tax lab	.27	

General Equilibrium (50% Subsidy)

	Edu. Participation (aggr.shares)			Human Capital Aggregates		
	Less than HS	HS	College	Less than HS	HS	College
	0.35	0.50	0.15	4.88	9.20	4.61
	Edu. Shares by ability			Marg. Products after Tax. and Depr.		
Ability 1 (lowest)	0.88	0.11	0.011	0.43	1.0	0.93
Ability 2	0.37	0.58	0.048	Aver. Post-Tax Labour Earn.-1990 US$		
Ability 3	0.25	0.56	0.19	556	1777	2361
Ability 4 (highest)	0.21	0.39	0.40			
% with wealth=0	0.081	r	0.025	Tax lab	.2734	

Partial Equilibrium (50% Subsidy)

	Edu. Participation (aggr.shares)			Human Capital Aggregates		
	Less than HS	HS	College	Less than HS	HS	College
	0.37	0.44	0.19	4.99	8.40	5.42
	Edu. Shares by ability			Marg. Products after Tax. and Depr.		
Ability 1 (lowest)	0.89	0.095	0.016	0.44	1.0	0.94
Ability 2	0.40	0.51	0.097	Aver. Post-Tax Labour Earn.-1990 US$		
Ability 3	0.27	0.49	0.24	589	1807	2257
group 4 (highest)	0.23	0.37	0.40			
% with wealth=0	0.084	r	0.025	Tax lab	.2750	

markets set-up. We use this model to evaluate alternative educational interventions.

In the current version we experiment with tuition subsidies. It becomes apparent that while in partial equilibrium such policies can be very effective in increasing education levels and reducing inequality in General Equilibrium the results are much less encouraging: the main effect of a subsidy there is to increase the supply of human capital as one would expect. However, it is the more able but liquidity constrained individuals who take up extra education, while the education levels of the less able can actually decrease (they are crowded out). Thus the subsidy acts on the composition of those in education.

In many respects this is in line with results found by Heckman, Lochner, and Taber (1998a). The inclusion of risky returns on labour earnings and the fact that labour supply is endogenous lend additional credibility to the result. The distributional changes in this economy under different interventions will be the focus of additional analysis. Moreover, future work includes assessing the relevance of liquidity constraints in the model economy and the equilibrium effects of artificially removing (insuring against) some of the risk components.

3.7 Appendix to chapter 3

3.7.1 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[46]. 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[47].

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

3.7.2 PSID 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.

[46] 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.

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

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.

The 1967-2000 Sample. 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 60 leaves us with 16,733 people. Dropping the self employment observations leaves 13,740 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 5559. Dropping individuals with unclear education records leaves 5,544 people in sample. Disposing of individuals with missing, top-coded or zero earnings reduces the sample to 5,112 individuals and dropping those with zero, missing or more than 5840 annual work hours brings the sample size to 5,102 individuals. We eliminate individuals with outlying earning records, defined as changes in log-earnings larger than 4 or less than -2, which leaves 4,891 individuals in the sample. Finally, dropping people connected with the SEO sample reduces the number of individuals to 2,791.

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

Table 3.7: Distribution of observations for the 1967-2000 sample, by year

year	Number of Observations	year	Number of Observations
1967	776	1983	1546
1968	842	1984	1582
1969	891	1985	1609
1970	952	1986	1632
1971	1069	1987	1624
1972	1168	1988	1631
1973	1250	1989	1639
1974	1290	1990	1600
1975	1342	1991	1628
1976	1385	1992	1564
1977	1442	1993	1551
1978	1466	1994	1486
1979	1502	1995	1437
1980	1535	1996	1363
1981	1512	1998	1293
1982	1505	2000	1191

Table 3.8: Distribution of observations for 1967-2000 sample, by education

years of education	Number of Individuals	Number of Observations
less than 12	364	5,358
12 to 15	1,621	25,358
16 or more	806	13,587

3.7.3 GMM estimation of the technology parameters

The Minimization Problem

After controlling for physical capital, the minimization problem we face in order to identify the technology parameters is the following

$$\min_{\{P\}_N \in \Re^N} \left\{ \widehat{F(H)} - \Phi\left(H_1, H_2, H_3, \{P\}_N\right) \right\}^2$$

where $\{P\}_N$ is a set of N parameters of the (possibly non-linear) function $\Phi(\cdot)$.

If we consider the case of a nested CES-CES function we can write the above problem as

$$\min_{\{A,B,r,s\} \in \Re^4} \left\{ \widehat{F(H)} - \left\{ AH_1^r + (1-A)\left[BH_2^s + (1-B)H_3^s\right]^{\frac{1}{s}} \right\}^{\frac{1}{r}} \right\}^2$$

Of course, depending on the procedure used to obtain $\widehat{F(H)}$, the residual term will be a different object.

To see this more clearly, consider a log linearisation of the problem above, such that we can write

$$\widehat{F(H)} \approx \left\{ AH_1^r + (1-A)\left[BH_2^s + (1-B)H_3^s\right]^{\frac{1}{s}} \right\}^{\frac{1}{r}} \exp^g$$
$$\log\left(\widehat{F(H)}\right) = \frac{1}{r}\log\left(\left\{ AH_1^r + (1-A)\left[BH_2^s + (1-B)H_3^s\right]^{\frac{1}{s}} \right\}\right) + g$$

where g is an error term capturing measurement error due to wage mis-reporting and errors in the approximation of the aggregate K.

Non linear method of moments (Minimum Distance Estimator)

Consider the original problem where we define the residual of our estimation as

$$\log\left(F(H)\right) - \frac{1}{r}\log\left(\left\{ AH_1^r + (1-A)\left[BH_2^s + (1-B)H_3^s\right]^{\frac{1}{s}} \right\}\right) = e$$

Of course there will be one such residual for each time period in the sample. We denote therefore a (column) vector of residuals with T elements as

$$[R]_{t=1}^{T}=\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix}$$

Two potential problems must be considered when minimizing the sum of such residual distances: (i) simultaneity in the determination of residuals and production inputs (human capital aggregates). This problem arises if error components (contained in the residual as defined above) also determine the employment decisions of agents in the economy. In this case we might expect a correlation between control variables and residuals which undermines the reliability of estimates of technology parameters[48]; (ii) The residuals, as defined above, might be characterized by a certain degree of autocorrelation over time which should be accounted for.

If none of the above mentioned problems was present, we could apply a very simple minimum sum of squares estimator, using the time vectors $\{H_{1t}, H_{2t}, H_{3t},\}$ as regressors. Denoting the transpose of a matrix X as X', we could write the simple non-linear minimization problem as

$$\min_{\{A,B,s,r\}} R'(A,B,s,r)\,\Omega^{-1}R(A,B,s,r)$$

where Ω is some (diagonal) weighting matrix used to account for possible heteroschedasticity of the residuals over time. In the homoschedastic case $\Omega=\sigma^2 I$ (the identity matrix).

If there was a problem of simultaneity in the determination of $\{H_{1t}, H_{2t}, H_{3t},\}$ and R_t the above method would not provide consistent estimates.

One way to control for the effects of simultaneity is to exploit orthogonality con-

[48]In this case $\{H_{1t}, H_{2t}, H_{3t},\}$ would be correlated with the residual dated t.

ditions that may hold between the residuals as defined above and some $L \times T$ matrix Z composed of L variables with T time observations per variable. We suppose that the number of variables L is sufficient to identify the parameters of the model, that is $L \geq N$ and we assume that Z is correlated with $\{H_{1t}, H_{2t}, H_{3t},\}$. The instruments' matrix is such that $E(R'Z) = 0$ and $E(\{H_1, H_2, H_3\}' Z) \neq 0$.

In general, we might have more IV's than parameters to estimate. In this case we cannot expect to satisfy the empirical counterpart of the population orthogonality conditions presented above because we have a system of $L > N$ equations

$$\widehat{R}'Z = m(P_N)$$

in only N unknowns. It is therefore reasonable to replace the unattainable requirement that $\widehat{R}'Z = 0$ with the requirement that $\widehat{R}'Z$ be small in some norm. Ignoring any multiplicative terms involving the sample size T, a candidate distance we might use as an objective function to minimize is

$$NORM = \widehat{R}'Z\Omega^{-1}Z'\widehat{R}$$

Hansen (1982) has shown that under some regularity assumptions, minimizing the NORM above produces a consistent estimator of the parameters P_N, and we can use any positive definite matrix Ω that is not a function of P_N.[49] The question is again what kind of weighting matrix Ω should be chosen. A natural way to proceed is to set Ω to the covariance matrix of the orthogonality conditions, that is

$$\Omega = COV\left(\widehat{R}'Z\right) = E\left\{Z'\widehat{R}\widehat{R}'Z\right\} = Z'E\left(\widehat{R}\widehat{R}'\right)Z$$

Unfortunately Ω is unknown and this adds to the estimation burden. However, if the covariance matrix can be written as $\Omega = \sigma^2\widetilde{\Omega}$ we can consider σ^2 an arbitrary constant,

[49]The general result is that if Ω is a positive definite matrix and if

$$p\lim \widehat{R}'(P_N)Z = 0$$

then the minimum distance (GMM) estimator of P_N is consistent.

rather than a separate unknown parameter: in fact, since $\widetilde{\Omega}$ is an unknown matrix, it can be arbitrarily scaled by some factor c, and if we rescale σ^2 by $\frac{1}{c}$ the product $\Omega = \sigma^2\widetilde{\Omega}$ remains the same. An example in which we could ignore σ^2 when minimizing the objective function is the classical case when $E\left(\widehat{R}\widehat{R}'\right) = \sigma^2 I$. This leads to the estimator

$$P_N = \arg\min \widehat{R}'Z\left(\sigma^2 Z'Z\right)^{-1} Z'\widehat{R} = \arg\min \widehat{R}'\wp_Z\widehat{R}$$

where $\wp_Z = Z\left(Z'Z\right)^{-1} Z'$ is the standard projection matrix in the $Z - space$. This is not different from a non-linear two stage least squares estimator, however it is more general in the sense that we are not limited to the above choice of Ω.

Any positive definite matrix Ω that is constant will deliver consistent estimates. However, efficiency of such estimates depend on the choice of the weigthing matrix Ω. Hansen has shown that $\Omega = Z'\Sigma Z$ where $\Sigma = E\left(RR'\right)$ is in fact an optimal choice.

When no time correlation is present we can therefore summarize the estimator matrix products as follows. The sample equivalent of the theoretical moment condition $E\left(R'Z\right) = 0$ is

$$\frac{1}{T}\widehat{R}'Z = \frac{1}{T}\sum_{i=1}^{T}\widehat{e}_i z_i' = 0$$

where $z_i' = \left(z_i^1, z_i^2, \ldots, z_i^L\right)$, so that the norm to minimize is

$$NORM = \frac{1}{T}\left(\sum_{i=1}^{T}\widehat{e}_i z_i'\right)\Omega^{-1}\frac{1}{T}\left(\sum_{i=1}^{T}\widehat{e}_i z_i\right)$$

The sample equivalent of the weighting matrix $\Omega = Z'\Sigma Z$ is the $(L \times L)$ White dispersion matrix, which is

$$\widehat{Z'\Sigma Z} = \frac{1}{T^2}\sum_{i=1}^{T} z_i z_i'\widehat{e}_i^2$$

and therefore we can express the objective function as

$$NORM = \frac{1}{T^2}\left(\sum_{i=1}^{T}\widehat{e}_i z_i'\right)\left(\frac{1}{T^2}\sum_{i=1}^{T} z_i z_i'\widehat{e}_i^2\right)^{-1}\left(\sum_{i=1}^{T}\widehat{e}_i z_i\right) =$$
$$\sum_{i=1}^{T}\widehat{e}_i z_i'\left(\sum_{i=1}^{T} z_i z_i'\widehat{e}_i^2\right)^{-1}\sum_{i=1}^{T}\widehat{e}_i z_i = \widehat{R}'Z\left(\widehat{Z'\Sigma Z}\right)^{-1} Z'\widehat{R}$$

For consistency of the estimates it is necessary that $\widehat{Z'\Sigma Z}$ is constant when minimizing the above NORM. Using $\widehat{Z'\Sigma Z} = I$ will deliver consistent but inefficient estimates. Estimation of any other $\widehat{Z'\Sigma Z}$ requires that some estimate of P_N is already in hand, even if P_N is the object of estimation: such estimate of P_N used to construct $\widehat{Z'\Sigma Z}$ may not be efficient but must be consistent in order to improve the efficiency of the main estimation procedure. This still leaves the open question of where to find the first round consistent estimator of P_N; one possibility is to obtain an inefficient but consistent GMM estimator by using $\widehat{Z'\Sigma Z} = I$ and then use the resulting estimator to construct $\widehat{\Sigma}$ which can be used to re-compute the NORM to minimize.

The GMM covariance matrix

Given the point estimates obtained from the minimization problem outlined before, we are interested in obtaining a (asymptotic) covariance matrix. Using a standard strategy, we can recover the asymptotic behavior of the estimator.

In general, ignoring the averaging factor $\frac{1}{T}$, the matrix $\Omega = Z'\Sigma Z$ is equal to

$$\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j' COV\left(\widehat{e}_i, \widehat{e}_j\right)$$

where z_i' is the $i-th$ row of Z , and if we denote z_h^l as the $h-th$ observation of instrument l we can rewrite this product as

$$Z'\Sigma Z = \sum_{i=1}^{T}\sum_{j=1}^{T} \begin{pmatrix} z_i^1 \\ z_i^2 \\ \vdots \\ z_i^L \end{pmatrix} \begin{pmatrix} z_j^1 & z_j^2 & \cdots & z_j^L \end{pmatrix} COV\left(\widehat{e}_i, \widehat{e}_j\right) =$$

$$\sum_{i=1}^{T}\sum_{j=1}^{T} \begin{pmatrix} z_i^1 z_j^1 & z_i^1 z_j^2 & \cdots & z_i^1 z_j^L \\ z_i^2 z_j^1 & z_i^2 z_j^2 & \cdots & z_i^2 z_j^L \\ \vdots & \cdots & \ddots & \vdots \\ z_i^L z_j^1 & \cdots & \cdots & z_i^L z_j^L \end{pmatrix} COV\left(\widehat{e}_i, \widehat{e}_j\right)$$

Assuming that this double summation divided by $\frac{1}{T^2}$ converges to a positive definite matrix, its estimation relies on the current estimates of the parameters P_N.If residuals are uncorrelated over time, the cross terms can be omitted as $COV(\widehat{e}_i, \widehat{e}_j) = 0$ when $i \neq j$ and we have that

$$Z'\Sigma Z = \sum_{i=1}^{T} z_i z_i' VAR(\widehat{e}_i)$$

which can be written in more extensive form as

$$Z'\Sigma Z = \sum_{i=1}^{T} \begin{pmatrix} z_i^1 \\ z_i^2 \\ \vdots \\ z_i^L \end{pmatrix} \begin{pmatrix} z_i^1 & z_i^2 & \cdots & z_i^L \end{pmatrix} VAR(\widehat{e}_i) =$$

$$\sum_{i=1}^{T} \begin{pmatrix} z_i^1 z_i^1 & z_i^1 z_i^2 & \cdots & z_i^1 z_i^L \\ z_i^2 z_i^1 & z_i^2 z_i^2 & \cdots & z_i^2 z_i^L \\ \vdots & \cdots & \ddots & \vdots \\ z_i^L z_i^1 & \cdots & \cdots & z_i^L z_i^L \end{pmatrix} VAR(\widehat{e}_i)$$

The White variance matrix estimator approximates this as

$$\widehat{Z'\Sigma Z} = \sum_{i=1}^{T} z_i z_i' \widehat{e}_i^2 = \sum_{i=1}^{T} \begin{pmatrix} z_i^1 z_i^1 & z_i^1 z_i^2 & \cdots & z_i^1 z_i^L \\ z_i^2 z_i^1 & z_i^2 z_i^2 & \cdots & z_i^2 z_i^L \\ \vdots & \cdots & \ddots & \vdots \\ z_i^L z_i^1 & \cdots & \cdots & z_i^L z_i^L \end{pmatrix} \widehat{e}_i^2$$

For the autocorrelation case, we can either use the Newey-West estimator of $Z'\Sigma Z$ or we can explicitly control for the presence of autocorrelation in residuals.

Testing

One of the additional benefits of the GMM testing method is that whenever the P_N is overidentified $(L > N)$ the minimand is also a test statistic for the validity of these restrictions. Under the null hypothesis that the overidentifying restrictions are valid it can be proven that

$$NORM = \sum_{i=1}^{T} \widehat{e}_i z_i' \left(\sum_{i=1}^{T} z_i z_i' \widehat{e}_i^2 \right)^{-1} \sum_{i=1}^{T} \widehat{e}_i z_i \sim^a \chi_2 (L - N)$$

This test does not however give any indication about the validity of all the instrumental variables, but answers the simpler question: given that a subset of the instrumental variables is valid and exactly identifies the coefficients, are the extra instrumental variables valid ?

Instruments' choice

In what follows we present some results obtained by applying the above method to the log-linearized version of the production function in which we set the elasticity parameters of the CES to zero (that is $r = s = 0$).

We find that a GMM procedure applied to the unrestricted CES specification provides poor, scarcely robust and highly insignificant estimates for all technology parameters. On the other hand, a restricted ($r = s = 0$) CES technology delivers a Cobb-Douglas specification of the form $F(H) = H_3^A \left(H_2^B H_1^{1-B} \right)^{1-A} \exp^f$ which can be easily log-linearized as

$$\ln F(H_t) = A \ln H_{3t} + (1 - A) [B \ln H_{2t} + (1 - B) \ln H_{3t}] + f_t$$

and given the small sample dimension (30 observations) this linearization makes the GMM procedure more robust and reliable. In fact, in a C-D specification it does not matter whether H_2 is nested with H_1or H_3. Such distinction would matter only in a CES specification.

In order to explicitly model possible error correlation we assume that

$$f_t = \theta f_{t-1} + \varepsilon_t$$
$$\varepsilon_t \quad \text{i.i.d.}$$

If we then denote $A\ln H_{3t}+(1-A)\left[B\ln H_{2t}+(1-B)\ln H_{3t}\right]$ as $X_t'\beta$, we can redefine the residuals to be used in computing the empirical moments as

$$\varepsilon_t=\ln F\left(H_t\right)-\rho\ln F\left(H_{t-1}\right)-X_t'\beta+\rho X_{t-1}'\beta \tag{3.16}$$

and by doing so we explicitly control for the time correlation of f_t.

We initially include a time polynomial of the form $t\left(time,\gamma\right)=c+\gamma_1 time_t+\gamma_2 time_t^2+\gamma_3 time_t^3$ in the conditional mean of $\ln F\left(H_t\right)$. However, after some initial testing we conclude that only the linear time trend can be robustly estimated in most of our model specifications, the other parameters in the time polynomial being insignificant and erratic. Therefore we have a final error term specification of the form

$$\varepsilon_t=\ln F\left(H_t\right)-X_t'\beta-\gamma time_t-\rho\left[\ln F\left(H_{t-1}\right)-X_{t-1}'\beta-\gamma time_{t-1}\right] \tag{3.17}$$

The instruments used to control for the simultaneity of ε_t and the endogenous human capital aggregates in H_t are lagged values of H_t itself. We present estimates based on empirical moments such as

$$\frac{1}{T}\sum_{i=1}^{T}\hat{\varepsilon}_t H_{t-1-m-g} \quad where \ \ m=1,...,M$$

$$g\in\{0,1,2\}$$

where $H_{t-1-m-g}=\left[H_{1,t-1-m-g}\ ,\ H_{2,t-1-m-g}\ ,\ H_{3,t-1-m-g}\right]$. Given this notation it follows that $3\left(M+1\right)$ is the number of moment conditions used in estimation. The index g indicates the minimum lag that is employed as an instrument (e.g., when $g=0$ we use a specification with instruments dated between $t-1$ and $t-M$).

The parameters to estimate in the final specification (3.17) are $\{\beta,\rho,\gamma\}$ where $\beta=(A,B)$. Different sets of instruments are alternatively used. We report estimates of these parameters under a set of moment restrictions which differ in the:

- choice of M;

- choice of g;
- choice of the first step weigting matrix, that is either the identity (I) or the instrument cross-product $(Z'Z)$;

To minimize the objective function we use a simplex method algorithm first proposed by Nelder and Meade (1965). This method has the advantage to check whether a candidate set of estimates is a real minimizing set by using a quadratic expansion in the neighborhood of such set and verifying that the minimum of such quadratic form corresponds to the minimum found by the Simplex.

The results of the GMM estimation procedure of the log-linearized C-D technology are reported in the following tables (standard errors in parenthesis). Notice that the total number of observations (T) available to compute the moments depends on the number and lenght of the lagged instruments and is equal to $30 - 1 - m - g$.

The final line of each table reports the objective function value (weighted sum of empirical moments): this is a test of overidentifying restrictions and is distributed as a $\chi^2_{3(M+1)-N}$ where $N = 4$ is the number of parameters to estimate.

Table 3.9 reports results obtained by using: (i) dependent variable measured from aggregate wage bills and physical capital augmented to account for residential wealth and (ii) a weighting matrix is an identity matrix.

Table 3.10 reports results based on the same dependent variable but with a first stage weighting given by the positive definite matrix $Z'Z$.

3.7.4 Definition of stationary measure

Stationary measure μ^*

Definition 1 *Let* $(X, \mathcal{F}(X), \psi_j)$ *be a measure space, where* $X = \Theta \times \Im \times Z \times \bar{A}$ *is the state space and* $\mathcal{F}(X)$ *the* σ*-algebra on* X*. In order to define a stationary measure* ψ_j *we need a transition function* $Q : X \times \mathcal{F}(X) \longrightarrow [0,1]$ *such that, for* $F \subset \mathcal{F}(X)$ *,* $\psi_j = Q(F, \psi_j)$.

In order to construct Q we define the following conditional probability $\gamma = \gamma(\pi(\cdot))$

Table 3.9: GMM technology estimates, identity weights

Dep. Var. Based on Wage Bills and Aug. K, First Step Weight Matrix: I									
	g = 0			g = 1			g = 2		
	$M=1$	$M=2$	$M=3$	$M=1$	$M=2$	$M=3$	$M=1$	$M=2$	$M=3$
A	**0.177**	**0.719**	**0.497**	**0.371**	**0.497**	**0.428**	**0.305**	**0.260**	**0.234**
	(0.615)	*(0.178)*	*(0.111)*	*(0.279)*	*(0.189)*	*(0.127)*	*(0.295)*	*(0.200)*	*(0.140)*
B	**1.058**	**-0.414**	**0.470**	**0.005**	**0.567**	**0.776**	**0.961**	**0.783**	**0.901**
	(0.722)	*(1.218)*	*(0.330)*	*(0.522)*	*(0.363)*	*(0.143)*	*(0.244)*	*(0.115)*	*(0.088)*
ρ	**0.948**	**0.958**	**0.951**	**0.975**	**0.950**	**0.954**	**0.936**	**0.954**	**0.938**
	(0.032)	*(0.010)*	*(0.012)*	*(0.017)*	*(0.015)*	*(0.011)*	*(0.018)*	*(0.014)*	*(0.013)*
γ	**0.022**	**0.049**	**0.043**	**0.051**	**0.040**	**0.034**	**0.032**	**0.036**	**0.036**
	(0.038)	*(0.010)*	*(0.007)*	*(0.012)*	*(0.008)*	*(0.006)*	*(0.009)*	*(0.006)*	*(0.005)*
T	**28**	**27**	**26**	**27**	**26**	**25**	**26**	**25**	**24**
func.	**4.067**	**4.171**	**13.676**	**0.593**	**11.683**	**14.456**	**3.425**	**6.077**	**8.603**
d.f.	**2**	**5**	**8**	**2**	**5**	**8**	**2**	**5**	**8**
$\chi^2_{(0.95)}$	**5.991**	**11.070**	**15.507**	**5.991**	**11.070**	**15.507**	**5.991**	**11.070**	**15.507**

Table 3.10: GMM technology estimates, non-identity weights

Dep. Var. Based on Wage Bills and Aug. K, First Step Weight Matrix: $Z'Z$									
	g = 0			g = 1			g = 2		
	$M=1$	$M=2$	$M=3$	$M=1$	$M=2$	$M=3$	$M=1$	$M=2$	$M=3$
A	**0.395**	**0.775**	**0.468**	**0.404**	**0.615**	**0.548**	**0.295**	**0.284**	**0.299**
	(0.552)	*(0.188)*	*(0.105)*	*(0.278)*	*(0.161)*	*(0.139)*	*(0.310)*	*(0.207)*	*(0.141)*
B	**0.964**	**-1.188**	**0.764**	**-0.029**	**1.420**	**0.770**	**1.042**	**0.790**	**0.850**
	(0.960)	*(2.106)*	*(0.253)*	*(0.552)*	*(0.300)*	*(0.188)*	*(0.246)*	*(0.123)*	*(0.010)*
ρ	**0.951**	**0.958**	**0.944**	**0.972**	**0.929**	**0.950**	**0.934**	**0.952**	**0.939**
	(0.021)	*(0.010)*	*(0.013)*	*(0.016)*	*(0.013)*	*(0.011)*	*(0.018)*	*(0.014)*	*(0.012)*
γ	**0.027**	**0.055**	**0.036**	**0.052**	**0.023**	**0.035**	**0.029**	**0.035**	**0.038**
	(0.034)	*(0.010)*	*(0.007)*	*(0.011)*	*(0.009)*	*(0.006)*	*(0.010)*	*(0.006)*	*(0.004)*
T	**28**	**27**	**26**	**27**	**26**	**25**	**26**	**25**	**24**
func.	**3.838**	**2.878**	**13.765**	**0.743**	**8.785**	**14.729**	**2.941**	**6.254**	**9.262**
d.f.	**2**	**5**	**8**	**2**	**5**	**8**	**2**	**5**	**8**
$\chi^2_{(0.95)}$	**5.991**	**11.070**	**15.507**	**5.991**	**11.070**	**15.507**	**5.991**	**11.070**	**15.507**

$$
\begin{aligned}
&\gamma\left[x, y \in F\right] = \Pr\left\{y \in F \mid x\right\} = \\
= &\int_Z \pi\{z_{j+1} \mid z_j\}\, I\left\{\left(\theta, e_{j+1}\left(x\right), z_{j+1}, a_{j+1}\left(x\right)\right) \in F\right\} dz_{j+1}
\end{aligned}
$$

which represents the fraction of agents transiting from $x = (\theta, e, z, a) \in X$ into $F \subset \digamma\left(X\right)$. $I\left(\cdot\right)$ is an indicator function.

We can then use $\gamma\left(\cdot\right)$ to define the stationary measure ψ_j^* as

$$
\psi_j^*\left(F\right) = Q\left(F, \psi_j^*\right) = \int_X \gamma\left[x, y \in F\right]\; d\psi_j^*\left(x\right)
$$

3.7.5 Analytical derivation of the market clearing condition

The budget constraint of a generic agent in chapter 3 can be described as

$$
\begin{aligned}
&c_j + a_{j+1} = \\
&= \left[1 + r\left(1 - \tau_k\right)\right]\left(a_j + q_j\right) + w_e \exp^{\epsilon_j} n_j(1 - \tau_{n^e})\left(1 - \mathrm{d}_j\right) - \left(D_e - T_e\right)\mathrm{d}_j
\end{aligned}
$$

Such expression can be simplified by using equation (3.8) to express q_j, so that we obtain

$$
\begin{aligned}
&c_j + a_{j+1} = \\
&\left[1 + r\left(1 - \tau_k\right)\right] a_j + w_e \exp^{\epsilon_j} n_j(1 - \tau_{n^e})\left(1 - \mathrm{d}_j\right) - \left(D_e - T_e\right)\mathrm{d}_j
\end{aligned}
\tag{3.18}
$$

where $a_1 = q_1$, with $q_j = 0$ and $a_{j+1} = a_{j+1}\left(x\right)$ for $j = 2, ..., \overline{j}$. Notice that $E\left(q_1\right)$ is also described in (3.8).

By integrating this expression using the population distribution $\mu\left(x, j\right)$ we obtain

$$\sum_{j=1}^{\bar{j}} \zeta_j \left[\int_X c_j(x)\, d\psi_j(x) + \int_X a_{j+1}(x)\, d\psi_j(x) \right] = \tag{3.19}$$
$$= (1 + r(1-\tau_k)) \sum_{j=1}^{\bar{j}} \zeta_j \int_{\bar{A}} a_j\, d\psi_j(a) +$$
$$+ \sum_{j=1}^{\bar{j}} \zeta_j \int_X w_e \exp^{\epsilon_j} n_j(x)(1-\tau_{n^e})(1-\mathrm{d}_j(x))\, d\psi_j(x) +$$
$$- \sum_{j=1}^{\bar{j}} \zeta_j \int_X D_e \mathrm{d}_j(x)\, d\psi_j(x) + \sum_{j=1}^{\bar{j}} \zeta_j \int_X T_e \mathrm{d}_j(x)\, d\psi_j(x)$$

Using the government budget constraint we can rewrite (3.19) as

$$G + \sum_{j=1}^{\bar{j}} \zeta_j \left[\int_X c_j(x)\, d\psi_j(x) + \int_X a_{j+1}(x)\, d\psi_j(x) \right] =$$
$$= (1+r) \sum_{j=1}^{\bar{j}} \zeta_j \int_{\bar{A}} a_j d\psi_j(a) + \sum_{j=1}^{\bar{j}} \zeta_j \int_X w_e \exp^{\epsilon_j^{(\theta,e,z)}} n_j(x)(1-\mathrm{d}_j(x))\, d\psi_j(x) +$$
$$- \sum_{j=1}^{\bar{j}} \zeta_j \int_X D_e \mathrm{d}_j(x)\, d\psi_j(x)$$

Let FX denote capital flows. Now use the following relationships

1. $\sum_j \zeta_j \int_{\bar{A}} a_j\, d\psi_j(a) = K(r) - FX(r)$, by definition;

2. $F_K = r + \delta$, by profit maximization;

3. $F(K,H) = F_K\ K + \sum_{j=1}^{\bar{j}} \zeta_j \int_X w_e \exp^{\epsilon_j} n_j(x)(1-\mathrm{d}_j(x))\, d\psi_j(x)$, because $F(K,H)$ is homogeneous of degree 1;

Using relationship (1) we can write the last equation as

$$G+\sum_{j=1}^{\bar{j}}\zeta_j\left[\int_X c_j(x)\,d\psi_j(x)+\int_X a_{j+1}(x)\,d\psi_j(x)\right]= \tag{3.20}$$

$$(1+r)(K-FX)+\sum_{j=1}^{\bar{j}}\zeta_j\int_X w_e\exp^{\epsilon_j(\theta,e,z)}n_j(x)(1-\mathrm{d}_j(x))\,d\psi_j(x)+$$

$$-\sum_{j=1}^{\bar{j}}\zeta_j\int_X D_e\mathrm{d}_j(x)\,d\psi_j(x)$$

Then, using relationships (2) and (3) we obtain

$$G+\sum_{j=1}^{\bar{j}}\zeta_j\left[\int_X c_j(x)\,d\psi_j(x)+\int_X a_{j+1}(x)\,d\psi_j(x)\right]= \tag{3.21}$$

$$F(H,K)+(1-\delta)K\ +(1+r)FX-\sum_j\zeta_j\int_X D_e\mathrm{d}_j(x)\,d\psi_j(x)$$

which, using again relationship (1) can be written as

$$G+\sum_{j=1}^{\bar{j}}\zeta_j\left[\int_X c_j(x)\,d\psi_j(x)+\int_X a_{j+1}(x)\,d\psi_j(x)\right]= \tag{3.22}$$

$$=F(H,K)+\sum_j\zeta_j\int_{\bar{A}}a_j\ d\psi_j(a)-\delta K\ -rFX-\sum_j\zeta_j\int_X D_e\mathrm{d}_j(x)\ d\psi_j(x)$$

This is exactly the goods market clearing equilibrium condition.

3.7.6 Computational method

We solve the agent's problem via backward recursion, starting from the last period of life. The optimal consumption and leisure decisions are found through an Euler equation method. Given the non-convexity of the value functions of students we do not rely only on local methods to solve for the zeroes of Euler equations: instead, we use also a grid method which allows to find all the intervals in which the Euler equations change sign. This method splits the state-space into small intervals and computes the sign of the Euler residual at the boundaries of such intervals. Whenever a change in the sign of the Euler residual takes place place within an interval, a non-linear equation solver is used to find the zero of the Euler equation within that interval.

Expectation with respect to labour shocks are computed using Gauss-Hermite integration.

We store optimal decisions and value functions at grid points but households choices are not restricted to coincide with these points. We do linear interpolation to evaluate choices which lie between points.

Once the optimal policy have been stored, we run simulations based on age cohorts of 10,000 individuals, starting from the youngest and moving further up the age ladder. When all age groups have been simulated we compute the aggregate states and the associated prices. We use a sup-norm criterion to check whether the new prices are close enough to the old ones. If they are close enough in the chosen sup-norm, we claim to have converged to an equilibrium, otherwise we start solving the problem all over again using a linear combination of old and new prices.

Chapter 4

Education and Crime over the Life Cycle

4.1 Introduction

This chapter 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 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.

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[1]. These numbers beg the question of whether U.S. policy makers are using a cost-effective mix of policies in the fight against crime.

For reasons similar to those highlighted in Heckman, Lochner, and Taber (1998b), the analysis of alternative policies to tackle crime benefits from the use of a dynamic

[1]The figure comes from Maguire and Pastore (1995) and is at the lower end of available estimates. Section 4.5 reports alternative estimates.

equilibrium framework[2]. 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.

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.

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[3]. 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

[2]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 (2006), Cunha, Heckman, and Navarro (2004)

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

marginally expands the stock of inmates and the associated expenditure. As a consequence, the proportional labour 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 labour 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 in efficiency implies the policy can be financed at a marginally lower labour 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.

4.2 Related literature

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) study jointly the effect of changes in market returns and in the expected punishment within a calibrated, dynamic general equilibrium model of crime choice which they use to account for fluctuations in the US property crime rate in the past twenty years. Their model is closest to the present one and, in fact, we use it as our starting point. They show that the model is remarkably effective in accounting for the evolution of the property crime rate in the last twenty-five years on the basis of changes in wage inequality, age distribution and expected punishment. Their analysis is positive and abstracts from educational choices and imperfect substitutability among workers with alternative education levels. Our focus is more normative. Furthemore, endogenising investment in education allows us to carry out a cost-benefit analysis of a richer set of alternative policies. Cozzi (2004) also uses a calibrated equilibrium model to investigate the extent to which differences in poverty and labour market opportunities can rationalize the higher crime rates among African-American males.

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.

4.3 The model

4.3.1 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 $\Lambda_j = \Pi_{s=0}^{j-1} \lambda_s$.[4]

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

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

equals one if the individual is in education at age j while the latter assumes value one for an agent engaged in crime at age j. They are zero otherwise. The felicity function is

$$u(c_j, i_j^s, d_j^c) = u(c_j) + \psi_j(\theta) i_j^s + \chi d_j^c = \frac{c_j^{1-\sigma}}{1-\sigma} + \psi_j(\theta) i_j^s + \chi d_j^c. \qquad (4.1)$$

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 $E = \{e_0, e_1, e_2\}$.[5] 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[6]. 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. For a victim being the target

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

[6]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.

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.

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 $1-\pi_u$.[7]. 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

$$h_j(\theta, e) = \exp(\theta + \xi_j(e)), \tag{4.2}$$

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

$$Q(H, K) = H^{1-\alpha} K^{\alpha}, \tag{4.3}$$

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

[7]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.

human capital stock H is the aggregate

$$H = [\alpha H_1^{\rho} + \beta H_2^{\rho} + (1-\alpha-\beta)H_3^{\rho}]^{\frac{1}{\rho}}. \tag{4.4}$$

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 $\bar{j}$ are distributed to the newborns in such a way that the cross-sectional distributions of wealth across deceased and newborn agents coincide[8].

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

[8]The details of the mechanism generating bequests are discussed in Section 4.5.

receive their labour income or transfers and decide how much to consume and save[9].

4.3.2 Recursive representation

The consumers' optimization problem admits a recursive representation (see 2 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" for agent i is given by

$$y_j^i(\theta, e, l, v) = \begin{cases} sub_e(a) - D_e & \text{if } i = s \\ (1 - t_l)(1 - v)\,w(e)\,h_j(\theta, e) & \text{if } i = nc \\ (1 - t_l)(1 - v)\,w(e)\,h_j(\theta, e) + \alpha\bar{y} & \text{if } i = na \\ \alpha\bar{y} & \text{if } i = a \\ 0 & \text{if } i = pr \\ p & \text{if } i = r. \end{cases} \tag{4.5}$$

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

[9] 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.

is

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

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

$$a' \geq \bar{a},\ a'_{\bar{j}} \geq 0,\ a_0 \text{ given} \tag{4.6}$$

The value function of a current student satisfies the Bellman equation

$$V_j^s(\theta, e_n, a) = \max_c u(c) + \psi_j(\theta) + \beta\lambda_j V_{j+1}^s(\theta, e_n, a') \tag{4.7}$$

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

$$V_j^s(\theta, e_n, a_j) = \max_c u(c) + \psi_j(\theta) + \beta\lambda_j \max\{V_{j+1}^s(\theta, e_{n+1}, a'), E_l V_{j+1}^w(\theta, e_{n+1}, a', l)\} \tag{4.8}$$

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

$$V_j^w(\theta, e, a, l) = \max\left\{V_j^c(\theta, e, a, l) + \chi, V_j^{nc}(\theta, e, a, l)\right\}. \tag{4.9}$$

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[10] is

$$V_j^a(\theta,e,a,l) = \frac{1}{\Lambda_j}\left[\bar{u}\sum_{s=0}^{\tau-1}\beta^s\Lambda_{s+j} + \beta^\tau\Lambda_{j+\tau}E_l V_{j+\tau}^w(\theta,e,a_{j+\tau},l)\right]. \qquad (4.10)$$

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

$$V_j^i(\theta,e,a,l) = E_v\left[\max_c u(c) + \beta\lambda_j E_l V_{j+1}^w(\theta,e,a',l')\right]. \qquad (4.11)$$

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

4.3.3 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 over discrete variables.

Definition 2 *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 $\left\{c_j^i, a_j'^i\right\} : S \times \Psi \to R$, 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:*

[10] 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.

1. *Given* $\{r, w_n\}$, $\{c_j^i(s,v), a_j^{\prime i}(s,v), i_j^s(s), d_j^c(s)\}$ *for* $i \neq pr$ *solve the set of problems (4.7)-(4.11) and* V_j^i *are the associated value functions. Moreover,* $a_j^{\prime pr}(s,v) = (1+r)a$.

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

$$r + \delta = F_K \tag{4.12}$$

and

$$w(e_n) = F_{H_n}. \tag{4.13}$$

3. *Factor and product markets clear*[11] *or*

$$H_n = \int_{J\times[S|E]} lh_j(\theta, e_n)\, d\mu_j^{na}(\theta, e_n, a, l) \tag{4.14}$$

and

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

4. *The government budget is balanced*

$$M + G + P + SUB = t_k \hat{r} K + t_l \int_{J\times S} w(e_s) lh_j(\theta, e_s)\, d\mu_j^{na}(s) \tag{4.15}$$

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

$$\pi_v = \left(\int_{J\times S} d\mu_j^w(s)\right)^{-1} \int_{J\times S} d\mu_j^c(s). \tag{4.16}$$

6. *The average disposable labor income of employed workers satisfies*

$$\bar{y} = \left(\int_{J\times S} d\mu_j^{na}(s)\right)^{-1} \int_{J\times S} w(e) lh_j(\theta, e_s)\, d\mu_j^{na}(s). \tag{4.17}$$

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

4.4 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.

4.4.1 Skill price variation, ability and wages

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 (4.1) 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.

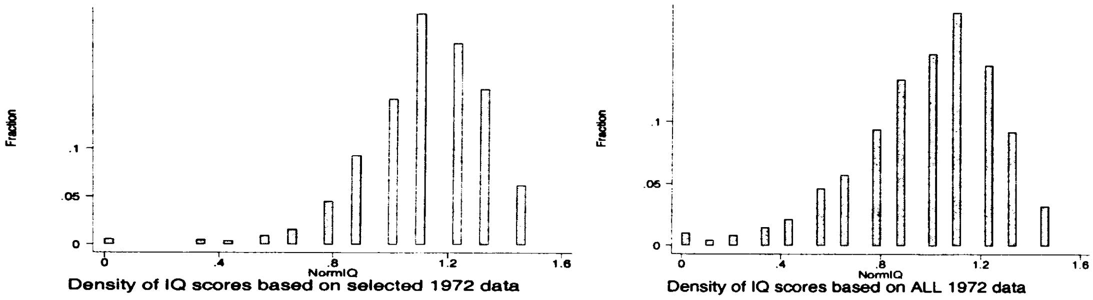

Figure 4.1: Density of IQ measurement from 1972 PSID wave, for the whole sample and a comparable sub-sample.

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

$$\ln w_{eit} = w_{et} + g_e\left(age_{it}\right) + u_{eit} \tag{4.18}$$

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

$$u_{eit} = b_e(\theta_i) + m_{it} \tag{4.19}$$

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

$$\ln w_{ei} = c^{edu} + \beta^{edu} IQ + \alpha_1 age + \alpha_2 age^2 + \varepsilon_{ei} \tag{4.20}$$

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[12].

In order to identify time variation in skill prices we exploit the panel dimension of the PSID data set. Using equation (4.18), we can identify the year-specific (log)

[12] 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.

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

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

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

$$\eta_{eit} = \widehat{residual}_{eit} - \widehat{residual}_{eit-1}$$

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

$$\eta_{ei,year} = dummy_{\{year\}} + \epsilon_{ei,year} \tag{4.21}$$

for all years between 1968 and 1997. Standard errors are robust and use cluster adjustment[13].

4.4.2 Estimating wage equations

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[14] and is

[13] No constant is estimated in this equation.

[14] 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.

described as total labor income of the head[15]. 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[16].

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 (4.1).

Table 4.1: Results for 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 (4.2) plots the age profiles (in logs) implied by the polynomial estimates for different education groups under the assumption that the constant terms are zero.

A detailed description of our sample selection for the estimation of log changes in

[15]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.

[16]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.

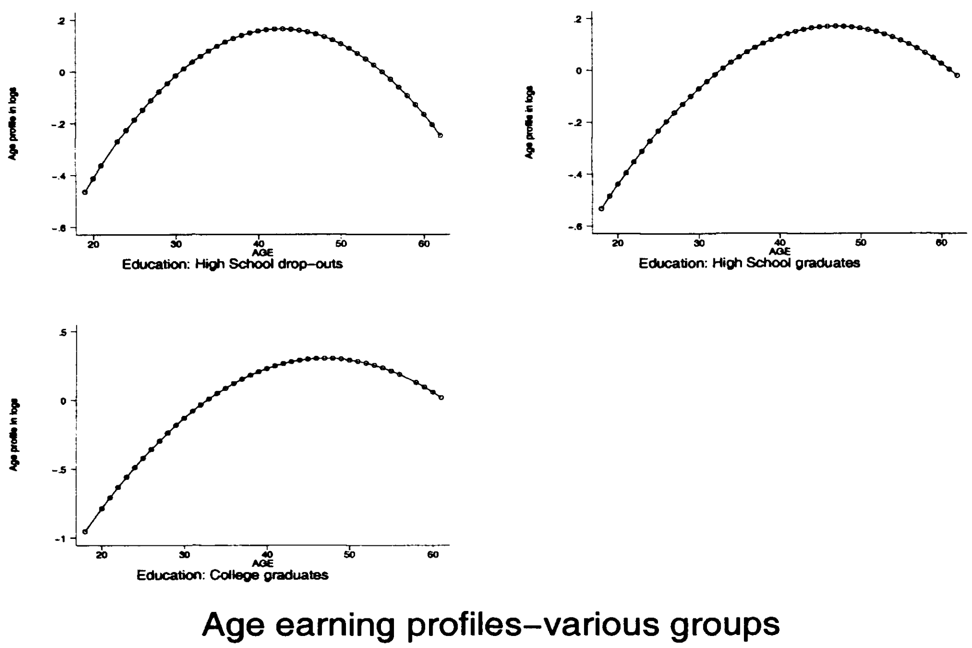

Figure 4.2: Log Age profiles of labour efficiency by education group

HC prices is reported in the Appendix to this chapter: 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 (4.3).

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

4.4.3 Aggregate production technology

We implement the method described in chapter 3 to measure human capital aggregates using CPS data.

However, unlike we did in chapter 3, we proceed to relax the restriction of unit elasticity implicit in the Cobb-Douglas specification and we only retain the isoelasticity

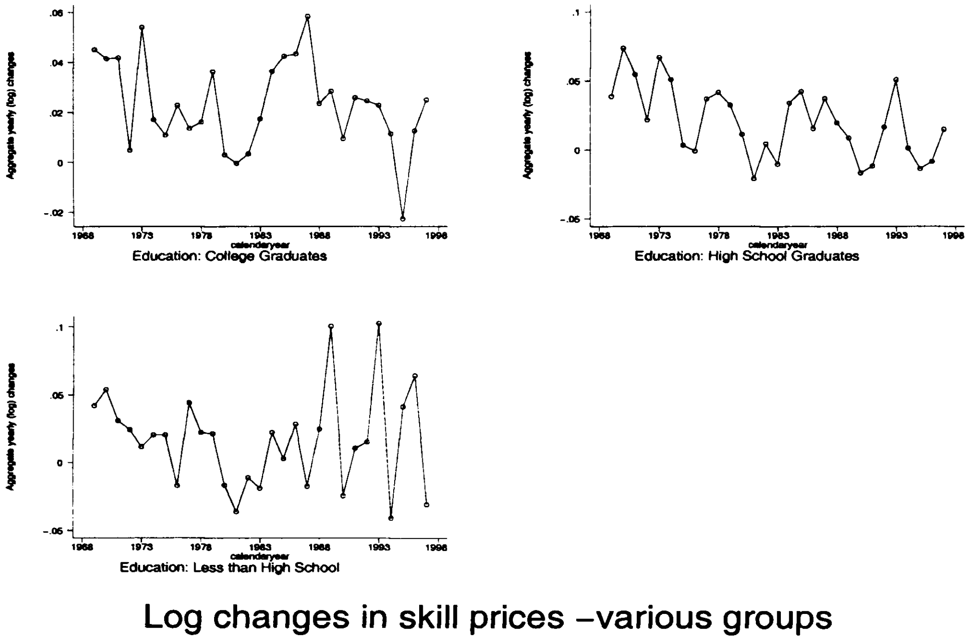

Figure 4.3: Log changes in marginal labor productivity, by education

assumption. We write the aggregate technology as

$$
\begin{aligned}
Y &= K^{\alpha} H^{1-\alpha} \\
H &= \left(aH_1^{\rho} + bH_2^{\rho} + (1-a-b)\, H_3^{\rho}\right)^{\frac{1}{\rho}}
\end{aligned}
$$

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

$$
\begin{aligned}
MPHC^3/MPHC^1 &= \frac{(1-a-b)}{a} \frac{H_3^{\rho-1}}{H_1^{\rho-1}} \\
MPHC^3/MPHC^2 &= \frac{(1-a-b)}{b} \frac{H_3^{\rho-1}}{H_2^{\rho-1}} \\
MPHC^2/MPHC^1 &= \frac{b}{a} \frac{H_2^{\rho-1}}{H_1^{\rho-1}}
\end{aligned}
$$

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

$$WB^3/WB^1 = \frac{(1-a-b)}{a}\left(\frac{H_3}{H_1}\right)^{\rho} \tag{4.22}$$

$$WB^3/WB^2 = \frac{(1-a-b)}{b}\left(\frac{H_3}{H_2}\right)^{\rho} \tag{4.23}$$

$$WB^2/WB^1 = \frac{b}{a}\left(\frac{H_2}{H_1}\right)^{\rho} \tag{4.24}$$

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 (4.20) 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 $\frac{H_2}{H_1} = \frac{\exp(\alpha_0^2)}{\exp(\alpha_0^1)}\frac{\tilde{H}_2}{\tilde{H}_1}$ 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

$$\log\left(WB^2/WB^1\right) = \left[\log\left(\frac{b}{a}\right) + \rho\log\frac{\exp\left(\alpha_0^2\right)}{\exp\left(\alpha_0^1\right)}\right] + \rho\log\left(\frac{\tilde{H}_2}{\tilde{H}_1}\right) \tag{4.25}$$

Notice that the ρ parameter in equation (4.25) 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

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

where (X, Y, Z) is a vector containing the estimates of the constant terms from the log-linearised estimation of equations $(4.22-4.24)$.

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

$$\delta_2-\delta_1+\rho\left(\alpha_0^2-\alpha_0^1\right)=X \tag{4.26}$$

$$\delta_3-\delta_1+\rho\left(\alpha_0^3-\alpha_0^1\right)=Y \tag{4.27}$$

$$\delta_3-\delta_2+\rho\left(\alpha_0^3-\alpha_0^2\right)=Z \tag{4.28}$$

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 $(\alpha_0^1, \alpha_0^2, \alpha_0^3)$ 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: 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.

4.4.4 Estimation results for production function parameters

We estimate a version of equation (4.25) 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 (4.20) 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 (4.4.4) with standard errors in parenthesis.

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 4.2: Unrestricted, log-linearised wage bills ratios equations.

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 4.3: Tests for equality of elasticities of substitution among HC inputs

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)[17]. The results of such tests are reported in table (4.4.4).

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 (4.20) in which we restrict the ρ to be the same for all ratios. The results for this specification are reported in table (4.4.4).

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

[17]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.

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 4.4: Restricted, log-linearised wage bills ratios equations

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[18]: 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 (4.25).

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 maximum working age,

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

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

Table 4.5: Value of intercepts in (log) age profile polynomial

the normalising assumption we are imposing is

$$\exp\left(\alpha_0^1\right) M^1 = \exp\left(\alpha_0^2\right) M^2 = \exp\left(\alpha_0^3\right) M^3 \tag{4.29}$$

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

$$\alpha_0^1 - \alpha_0^2 = \ln\left(M^2\right) - \ln\left(M^1\right) \tag{4.30}$$

$$\alpha_0^2 - \alpha_0^2 = \ln\left(M^3\right) - \ln\left(M^2\right) \tag{4.31}$$

which can be solved by setting the value of one of the α_0^{edu} terms to some arbitrary value and solving equations (4.30 – 4.31) for the remaining α_0^{edu} terms[19]. The resulting values are reported in table (4.4.4).

We plug these values from table (4.4.4) in the system (4.26 – 4.28) together with the constants estimated in the wage bills equations (X=.5943174,Y=-.2503577,Z=-.8446752) reported in table (4.4.4). 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}

[19]The solution of this system is:

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

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

Table 4.6: 1972 normalised human capital

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

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 (4.20). This allows to identify also a series of human capital aggregates which are consistent with the technology parameters. For clarity we report in table (4.6) both the constant terms we have estimated for equations (4.20) and the normalising α_0^{edu} obtained from equations $(4.30 - 4.31)$. 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 (4.21) 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 (4.4).

The price pattern reported in (4.4) 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 labour force and wage bills. 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

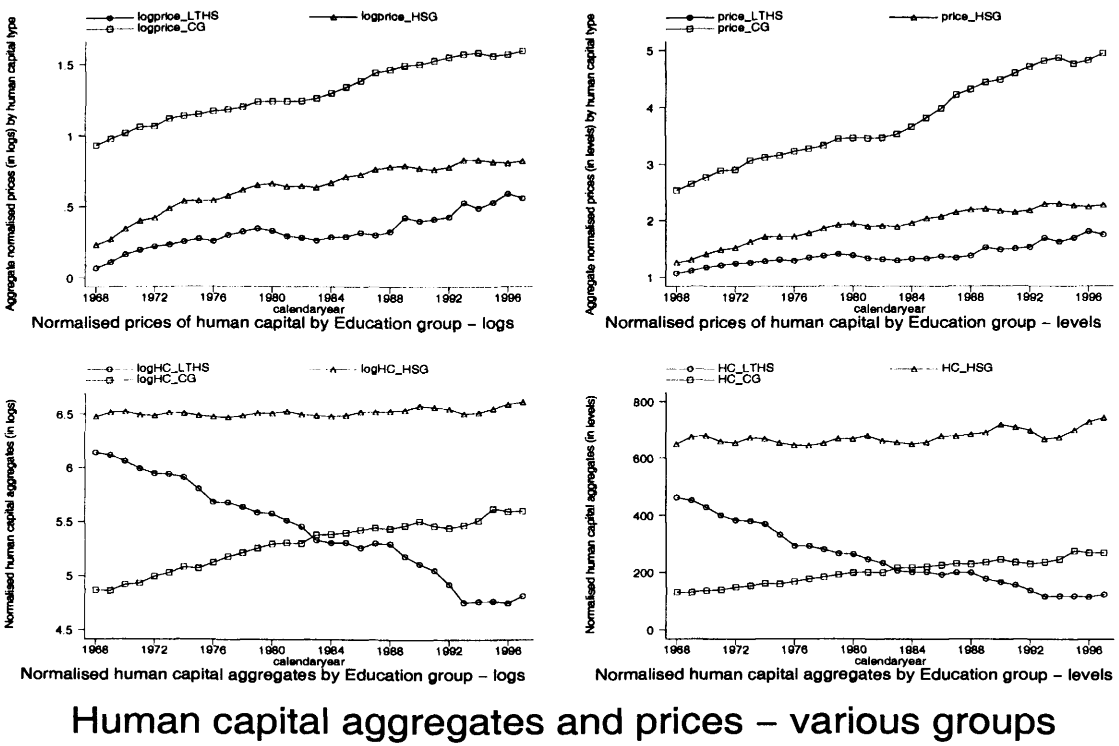

Figure 4.4: Human capital aggregates and associated prices, logs and levels

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 (3.8) for the Cobb-Douglas case.

4.5 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.[20]

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 (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 4.1. 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

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

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 4.7: Education shares among workers, by ability

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 outgoings. There does not seem to be not much information on such costs.

Education Enrolment Rates. Education 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 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 (4.7).

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

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 4.8: Grossed-up education shares among workers, by ability

period[21] which are reported in table (4.8.2). 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[22]. The values of the grossed-up education shares by ability are reported in table (4.8).

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%.[23] Furthermore, the evidence linking increased punishment to aggregate crime rate indicates that the elasticity of property crime with 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

[21] 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.

[22] 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.

[23] 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.

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 (4.35) with the exception of the utility quasilinear terms associated with the education decisions which are reported in table (4.8.2). Both tables are in the appendix.

4.6 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.

4.6.1 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 (4.8). 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 4.10), 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[24] which introduces selection based on wealth. In these bins, agents who opt to continue schooling at ages 16 are richer overall than those who opt for work. As shown in table 4.9, 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

[24] 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.

this bin high school dropouts account for roughly 30% of the population, see table 4.8). 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[25].

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.[26]

Average earnings and Income by Education. The model is able to replicate some stylised facts about inequality in earnings and income. Table (4.10) reports post-tax wage and income, together with average ability and consumption, 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

[25]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.

[26]See table (4.1) in the empirical section for an explanation of the zero gradient of IQ on college wages.

Average wealth of workers and students at ages 16 and 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 4.9: Average asset holding of young adults - simulated values in model units

Summary statistics for benchmark economy: averages						
	Education					
	HS dropouts		HS Graduates		College Graduates	
	Crime	No Crime	Crime	No Crime	Crime	No Crime
Ability	0.948	1.023	1.127	1.160	1.111	1.137
Wage	1.19	1.72	1.28	1.85	1.37	2.71
Ability by edu	1.011		1.159		1.137	
Wage by edu	1.63		1.84		2.70	
Income by edu	1.30		2.03		3.07	

Table 4.10: Summary statistics: average ability, income and wage. Benchmark

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 (4.11) we report some data on education and crime[27] based on the NLSY, provided by Lochner (2004).

[27]Thefts of at least $ 50 or shoplifting.

Self-reported criminal participation rates by education status Source: NLSY (males ages 20-23,1980), 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 4.11: Criminal participation rates by education group. S.E. in parenthesis

Crime rates by ability and education. 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 4.12: Simulated crime rates by ability and education. Benchmark

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.[28] 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 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 (4.12) 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

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

Crime rates by ability and education - 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 4.13: Changes in crime rates w.r.t. benchmark, by ability and education. Jail term: 13.2 months, G.E.

expenditure is equal to 0.43, that is 0.3% of aggregate output and 1% of total tax revenues.

4.6.2 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.

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 (4.13), 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 4.14 reports the new values. In the lowest ability group the

Changes in education shares by ability. 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 4.14: Changes in enrolment rates w.r.t. benchmark, by education and ability. Jail term: 13.2 months, G.E.

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%. 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%.

4.6.3 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[29]. 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 was run. Since students are likely to have benefited from the teaching support as well as the transfer,

[29]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.

Changes in education shares by ability. 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 4.15: Changes in enrolment rates w.r.t. benchmark, by education and ability. Non means-tested High School subsidy. P.E.

Changes in crime rates by ability and education. Non M.T. HS subsidy. P.E.

	Education group		
ability bin	*HS dropouts*	*HS 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 4.16: Changes in crime rates w.r.t. benchmark, by ability and education. Non means-tested High School subsidy. P.E.

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 4.15.

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 compensate the costs of the program. Aggregate prison costs plus the costs of the tuition programme are just 1% of total tax revenues. However, as shown in table 4.16, the positive effect in terms of crime associated to moving people away from the pool of high school dropouts 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 among different education groups.

In table 4.17 we report percentage changes (with respect to the benchmark) in

Differences (%) w.r.t. benchmark averages.	**Non M.T. HS subsidy. P.E.**	
	HS grads *vs* HS dropouts	College grads *vs* HS grads
Ability	-6.1%	0.0%
Labor price	0.0%	0.0%
Wage	-4.8%	1.2%
Income	54.8%	1.0%

Table 4.17: Change (%) w.r.t. benchmark in average ability, wage, income and labour price. Non means-tested HS subsidy. P.E.

between-group differences in average wage, ability, marginal returns[30] 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 4.18 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 4.19 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

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

Changes in crime rates by ability and education. Non M.T. HS subsidy. G.E.			
	Education group		
ability bin	*HS dropouts*	*HS 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 4.18: Changes in crime rates w.r.t. benchmark, by ability and education. Non means-tested High School subsidy. G.E.

Changes in education shares by ability. Non M.T. 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 4.19: Changes in enrolment rates w.r.t. benchmark, by education and ability. Non-means tested High School subsidy. G.E.

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 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 4.20, 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

Differences (%) w.r.t. benchmark averages. Non M.T. HS subsidy. G.E.		
	HS grads *vs* HS Dropouts	College grads *vs* HS graduates
Average ability	12.2%	31.8%
Labor price	-16.7%	-1.4%
Average wage	-4.8%	-1.2%
Average income	-5.5%	-2.9%

Table 4.20: Change (%) w.r.t. benchmark in average ability, wage, income and labour price. Non means-tested High School subsidy. G.E.

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 27.03%) as in the case of longer prison term. 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.

4.6.4 Restricting the target population

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 do not change the

Changes in education shares by ability. M.T. 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 4.21: Changes in enrolment rates w.r.t. benchmark, by education and ability. Means-tested High School subsidy. P.E.

Changes in crime rates by ability and education. M.T. HS subsidy. P.E.			
	Education group		
ability bin	*HS dropouts*	*HS 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 4.22: Changes in crime rates w.r.t. benchmark, by ability and education. Means-tested High School subsidy. P.E.

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 (4.21) 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 4.22). This curious fact can be rationalised by looking at table (4.23), 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 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

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

Table 4.23: Share of eligible agents taking up subsidy. Means-tested HS subsidy experiment in P.E.

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[31]. 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 (4.24) 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 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.

[31]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 *vs* non-takers/non-eligible. M.T. 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 4.24: Takers vs non-eligible or non takers. Average wealth endowment at time of choice (age 16), average ability and average crime rate over life cycle. Means-tested High School subsidy. P.E.

Changes in education shares by ability. M.T. 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 4.25: Changes in enrolment rates w.r.t. benchmark, by ability. Means-tested High School subsidy. G.E.

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.

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 (4.25 – 4.28) 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 difference is larger: price effects adjust in such a way

Changes in crime rates by ability and education. M.T. HS subsidy. 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 4.26: Changes in crime rates w.r.t. benchmark, by ability and education. Means-tested High School subsidy. G.E.

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

Table 4.27: Share of eligible agents taking up subsidy. Means-tested High School subsidy. G.E.

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%. 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

Eligible takers *vs* non-takers/non-eligible. M.T. 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 4.28: Takers vs non-eligible or non takers. Average wealth endowment at time of choice (age 16), average ability and average crime rate over life cycle. Means-tested High School subsidy. G.E.

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.

4.6.5 Introducing a control group: A 50/50 randomisation

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 do not 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 (4.29).

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 4.6.3. 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

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 4.29: 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.

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.

4.7 Conclusions

In this chapter 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.

4.8 Appendix to chapter 4

4.8.1 PSID sample selection

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 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.

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 4.30: 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 4.31: Distribution of individuals and total observations by years of education.

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 4.32: 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.

4.8.2 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. 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 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.[32] 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.

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.[33] 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 (4.8.2), categorized by education group and year.

[32]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.

[33]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.

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 4.33: 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.

year	Average Earnings (yearly, in $) censored data			adjusted data			year	Average Earnings (yearly, in $) 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 4.34: 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 4.35: 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 4.36: 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.

4.8.3 Numerical implementation

The procedure adopted to implement the numerical simulations is very similar to the one described in section 3.7.6. However, the inclusion of an additional binary decision makes the execution much slower.

Chapter 5

Conclusion

This thesis focussed on the evaluation of policy interventions adopting a structural approach. Chapter 2 has dealt with issues stemming from the inclusion of binary decisions within life-cycle models and has provided a set of results that are useful to understand the characteristics of such models and to implement their numerical counterparts. Chapter 3 has looked at how subsidization of College education interacts with heterogeneity, addressing issues of selection and wage inequality in equilibrium. Chapter 4 has provided an example of structural policy analysis comparing the effectiveness of alternative interventions targeting a reduction in crime rates.

The analysis in chapters 3 and 4 has relied on the measurement of fundamental model parameters through direct estimation or model calibration. Aggregate moments were mostly calibrated whereas individual parameters were estimated. Estimation equations were specified in accordance to the models' structure. Different approaches to the estimation of wage dynamics and production technologies have been presented in different chapters.

Through counterfactual policy analysis the chapters of this thesis have provided evidence supporting the view that long-term, equilibrium policy outcomes may drastically differ from their short-term counterparts. The comparison of partial and general equilibrium effects showed that price responses, even of relatively small size, can trigger large changes in selection into (or out of) education and crime. In chapter 3 it was shown that College subsidization could actually lead to increased wage inequality when all equilibrium effects are accounted for and that major crowding out effects can take place in equilibrium. Chapter 4, similarly, has documented the importance of earnings and wealth inequality in shaping criminal behaviour and showed that policies which are able to induce reductions in equilibrium wage inequality turn out to be extremely effective in reducing crime participation. In all applications equilibrium effects have proved to be non trivial, both in quality and quantity.

These results have highlighted the significance of explicitly modelling different sources of heterogeneity in the evaluation of policy interventions: the distributional effects of policies appear to be remarkably strong and able to shape aggregate outcomes. Some extensions could make these models more interesting and are planned as future research. First, comparing measures of welfare would identify the winners and losers of different interventions. This would help understand whether a specific intervention is implementable from a political economic perspective. Second, the computation of transitional dynamics could shed some light on the effects of a policy intervention while agents adjust to the new steady state. This would complement the welfare analysis, by indicating how painful a transition really is. The future research agenda should also include the study of alternative methods to parameterise large heterogeneous agents models in a consistent manner. Better procedures to reconcile the agent-based, behavioural dimension with the aggregate dimension would be an enormous step forward.

Bibliography

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

ACEMOGLU, D. (2002): "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, 40(1), 7–72.

ALONSO-BORREGO, C., J. FERNANDEZ-VILLAVERDE, AND J. GALDON-SANCHEZ (2004): "Evaluating Labor Market Reforms: A General Equilibrium Approach," Economics Working Papers we042307, Universidad Carlos III, Departamento de Economa, available at http://ideas.repec.org/p/cte/werepe/we042307.html.

ANGELUCCI, M., AND G. DEGIORGI (2006): "Indirect Effects of an Aid Program: The Case of Progresa and Consumption," Mimeo.

ATTANASIO, O., AND G. WEBER (1993): "Consumption Growth, the Interest Rate and Aggregation," *Review of Economic Studies*, 60(3), 631–649.

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

BECKER, G. S. (1964): *Human Capital.* NBER.

BECKER, R., AND J. BOYD (1997): *Capital Theory, Equilibrium Analysis and Recursive Utility.* Harvard University Press, Cambridge, MA.

BEN-PORATH, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), 352–365.

BERGE, C. (1997): *Topological Spaces*. Dover, New York.

BLACKORBY, C., AND R. R. RUSSELL (1989): "Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities)," *American Economic Review*, 79(4), 882–88.

BLINDER, A. S., AND Y. WEISS (1976): "Human Capital and Labor Supply: A Synthesis," *Journal of Political Economy*, 84(3), 449–472.

BLUNDELL, R., M. BROWNING, AND C. MEGHIR (1994): "Consumer Demand and the Life-Cycle Allocation of Household Expenditure," *Review of Economic Studies*, 61(1), 57–80.

CASELLI, F., AND W. J. COLEMAN (2006): "The World Technology Frontier," *American Economic Review*, 96(3).

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.

——— (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57(2), 191–261.

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.

EATON, J., AND H. S. ROSEN (1980): "Taxation, Human Capital and Uncertainty," *American Economic Review*, 70(4), 705–715.

FERNANDEZ, R., AND R. ROGERSON (1995): "On the Political Economy of Education Subsidy," *Review of Economic Studies*, 62, 249–262.

——— (1998): "Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform," *American Economic Review*, 88(4), 813–833.

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.

FRENCH, E. (2000): "The Labor Supply Response To (Mismeasured but) Predictable Wage Changes," Mimeo, Federal Reserve Bank of Chicago.

FRIEDMAN, M. (1957): *A theory of the consumption function.* Princeton University Press, Princeton.

GALE, W., AND J. SCHOLZ (1994): "Intergenerational Transfers and the Accumulation of Wealth," *Journal of Economic Perspectives*, 8(4), 145–160.

GOMES, J., J. GREENWOOD, AND S. REBELO (2001): "Equilibrium Unemployment," *Journal of Monetary Economics*, 48, 109–152.

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.

HAMERMESH, D. S. (1993): *Labor Demand.* Princeton, N. J.: Princeton University Press.

HANSEN, L. P. (1982): "Large Sample Moments of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HEATHCOTE, J., K. STORESLETTEN, AND G. VIOLANTE (2003): "The Macroeconomic Implications of Rising Wage Inequality in the US," Mimeo.

HECKMAN, J. (1981): "The Incidental Parameters Problem and the Problem of Initial Condition in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski, and D. L. McFadden. MIT Press.

——— (1992): "Randomization and Social Policy Evaluation," in *Evaluating Welfare and Training Programs*, ed. by C. F. Manski, and I. Gerfinkel. Cambridge: Harvard University Press.

HECKMAN, J., AND P. CARNEIRO (2003): "Human Capital Policy," NBER Working Papers 9055, National Bureau of Economic Research, Inc.

HECKMAN, J., L. LOCHNER, AND R. COSSA (2002): "Learning-By-Doing Vs. On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation," NBER Working Papers 9083, National Bureau of Economic Research, Inc.

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

——— (1998b): "General Equilibrium Treatment Effects: A Study of Tuition Policy," *American Economic Review, Papers and Proceedings*, 88(2), 381–386.

——— (1998c): "Tax Policy and Human Capital Formation," *American Economic Review, Papers and Proceedings*, 88(2), 293–297.

HECKMAN, J. J. (1976): "A Life-Cycle Model of Earnings, Learning, and Consumption," *Journal of Political Economy*, 84(4), S11–S44, Part 2: Essays in Labor Economics in Honor of H. Gregg Lewis.

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

HYSLOP, D. (2001): "Rising U.S. Earnings Inequality and Family Labor Supply: the Covariance Structure of Intrafamily Earnings," *American Economic Review*, 91(4), 755–777.

IMROHOROGLU, A., A. MERLO, AND P. RUPERT (2004): "What Accounts for the Decline in Crime," *International Economic Review*, 45, 707–729.

JEROME, A., C. DUSTMANN, C. MEGHIR, AND J. M. ROBIN (2006): "Career Progression and Formal versus on the Job Training," Mimeo.

KEANE, M., AND K. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

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 (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 74(1), 1–46.

LEVHARI, D., AND Y. WEISS (1974): "The effect of risk on the investment in human capital," *American Economic Review*, 64, 950–63.

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).

LUCAS, R. E. (1980): "Equilibria in a Pure Currency Economy," in *Models of Monetary Economics*, ed. by J. Kareken, and N. Wallace. Federal Reserve Bank of Minneapolis, Minneapolis.

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.

MCFADDEN, D. (1963): "Constant Elasticity of Substitution Production Functions," *Review of Economic Studies*, 30(2), 78–83.

MEGHIR, C. (2006): "Dynamic models for policy evaluation," IFS Working Papers W06/08, Institute for Fiscal Studies, available at http://ideas.repec.org/p/ifs/ifsewp/06-08.html.

MEGHIR, C., AND L. PISTAFERRI (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

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.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66(4), 281–302.

MINCER, J. (1991): "Human Capital, Technology and the Wage Structure, What Do Time Series Show?," Mimeo.

——— (1994): "Investment in U.S. Education and Training," Discussion Paper NBER WP4844, NBER, Working Paper Series.

MODIGLIANI, F., AND R. BRUMBERG (1954): "Utility analysis and the consumption function: an interpretation of cross-section data," in *Post-Keynesian Economics*, ed. by K. Kurihara. Rutgers University Press, New Brunswick.

NELDER, J. A., AND R. MEADE (1965): "The Role of Pre-market Factors in Black-White Wage Differences," *Computer Journal*, 7(5), 308–313.

PAVONI, N. (2006): "Optimal Unemployment Insurance with Human Capital Depreciation and Duration Dependence," Mimeo, UCL.

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.

ROSEN, S. (1977): "Human Capital: Relations between Education and Earnings," in *Frontiers of Quantitative Economics*, ed. by M. D. Intriligator, vol. 3b. Amsterdam: North Holland.

SATO, K. (1967): "A Two-Level Constant-Elasticity-of-Substitution Production Function," *Review of Economic Studies*, 34(2), 201–218.

STINCHCOMBE, M. B., AND H. WHITE (1992): "Some Measurability Results for Extrema of Random Functions over Random Sets," *Review of Economic Studies*, 59(3), 495–514.

STOKEY, N. L., R. E. LUCAS, AND E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press.

STORESLETTEN, K., C. I. TELMER, AND A. YARON (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51(3), 609–633.

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

TOWNSEND, R., AND K. UEDA (2001): "Transitional Growth with Increasing Inequality and Financial deepening," mimeo, University of Chicago.

TOWNSEND, R. M., AND K. UEDA (2003): "Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation," Discussion Paper 238, International Monetary Fund, IMF Working Papers 03/193.

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

WILLIS, R., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5), s7–s36, pt. 2.

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

YAARI, M. E. (1965): "Uncertain Lifetime, Life Insurance and the Theory of the Consumer," *Review of Economic Studies*, 32(2), 137–150.