Human Capital Accumulation, Education Policy and Wages Dispersion

Giovanni Gallipoli
Costas Meghir
Gianluca Violante*†

July 9, 2004

(PRELIMINARY AND INCOMPLETE)

Abstract

This paper examines the effectiveness of alternative policies aimed at changing the distribution of education in both partial and general equilibrium. Empirical evidence suggests a link between human capital (HC) accumulation and wages dispersion, so that policies affecting education outcomes will also have an impact on inequality, productivity and welfare. We build a life-cycle model of labor earnings and endogenous education choice, allowing for agents' heterogeneity in several dimensions. PSID and CPS data are used to estimate the parameters of a production function with different kinds of human capital and the transition laws of idiosyncratic labor shocks. A by-product of the estimation procedure is an empirical density of individual permanent efficiency over the working population, inferred from the empirical wages distribution. These estimates are used to pin down some of the model's parameters. Our benchmark model is able to replicate recent levels of US earnings inequality and education enrolment rates. Numerical simulations are used to compare the effects of alternative policy interventions on education participation, endogenous selection, life-cycle earnings, wealth profiles, inequality and productivity.

*University College London and IFS/CEMMAP; University College London and IFS; New York University.

†Special thanks go to Lars Nesheim for many discussions on both theoretical and computational aspects of the problem. We are also grateful to Orazio Attanasio, Marco Bassetto, Michele Boldrin, Martin Browning, Monica Costa Dias, Marco Cozzi, Maria Cristina DeNardi, Hide Ichimura, Ken Judd, Hamish Low, Robert Moffitt, Nicola Pavoni, Josep Pijoan-Mas, Jean-Marc Robin, Victor Rios-Rull, Robert Townsend and Fabrizio Zilibotti for comments and suggestions. All errors are ours.

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

1 Introduction

This paper examines human capital policies designed to alter the equilibrium distribution of education and their wider economic consequences. It also looks at the nature of education decisions and the role that such decisions play in shaping life cycle earnings and wealth profiles. Individual choices are analyzed in the context of a general equilibrium model with a spot job-market for each education group such that the unit price of (efficiency-weighted) labor depends on an agent's education and corresponds to its marginal product.

We assess the effectiveness of policy interventions targeting the wider population rather than limited groups, with relative labor prices endogenously adjusting to changes in the aggregate supply of educated people[1]. We analyze existing policy instruments, such as tuition transfers and loan subsidies[2], but we also devise and evaluate 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 tuned to match specific long-term features of the US labor market. By simulating and comparing equilibrium outcomes we aim to explore the quantitative aspects of the relationship among human capital accumulation, wages inequality and education policy. The impact of diverse education policies on equilibrium measures of productivity, consumption and welfare is also analyzed.

Research linking HC 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 (1994). Studies on the evaluation of policy interventions are more recent. In a key contribution to the empirical literature on education policy Keane and Wolpin (1997) study the partial equilibrium effect of a tuition subsidy on young males' college participation, while Donghoon Lee (2001) generalizes their approach to general equilibrium. In related work,

[1]Admittedly, given that labor is bought and sold on spot markets,the demand for labor is always equal to the supply. Alternatively, Acemoglu (1998) 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.

[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, altring HC investment decisions.

Heckman, Lochner and Taber (1998) estimate and simulate a dynamic general equilibrium model of education accumulation, assets accumulation and labor earnings with skill-biased technological change. 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 (see Blinder and Weiss, JPE 1984).

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 transitory efficiency shock[3].

Each agent in our model represents a household, which is intended as a family unit with possibly more than one individual supplying labor. Recent empirical evidence (Hyslop, 2001) indicates that labor supply explains 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[4]. 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 (1980a) 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 given levels, it is the case that such levels differ substantially 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 [5]. Labor supply, therefore, represents an effective channel of adjustment to labor price signals and an important determinant of the relative variations in skill prices[6].

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 David Levhari and Yoram 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 indi-

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

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

[5]To this purpose we include a state variable capturing permanent unobserved characteristics in our model.

[6]An example of the importance of differences among education groups in life time labor supplies come from earnings taxation. Taxes on labor earnings 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.

vidual 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[7] Eaton and Rosen (1980a) 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 to human capital investment[8]; 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.

We consider three levels of education obtained through formal schooling and corresponding to three types of HC separately entering the production technology[9]. Education and employment are mutually exclusive in each period. Foregone earnings and tuition charges are the direct costs of schooling, and a utility cost comes in the form of reductions in leisure when studying.

Agents can accumulate real assets and we experiment with alternative ways to assign assets to new borns[10]. We also consider different levels of correlation between ability and initial assets holdings[11].

In general, the model provides a way to look at endogenous equilibrium levels of aggregate human capital, with associated wages, as a function of agents' optimising schooling choices and demographic factors; furthermore, it represents a mapping from the initial agents' distributions (that is, their 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 implications of different education distributions.

2 Model[12]

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 used as physical capital. Different kinds of

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

[8]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.

[9]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.

[10]Of special interest is the case when the initial distribution of wealth replicates that prevailing among people who died in the previous period, as this has a realistic accidental bequest interpretation.

[11]This can be thought as a shortcut to incorporate the effect of parental background on ability formation, as extensively documented in the literature (see Heckman and Carneiro, 2002, for a review).

[12]A detailed illustration of this (and similar) models can be found in Gallipoli (2004). This includes a discussion of the implications of non-convexities of budget sets for existence and uniqueness of the solutions.

human capital are an input to the aggregate production function and command different returns. Wage differences among people are the consequence of differences in education (between group inequality) and differences in labor efficiency (within group inequality). We assume that people with different labor efficiencies are perfect substitutes within schooling groups. Agents can accumulate assets representing ownership of physical capital.

2.1 Demographics and preferences

Each (non-altruistic) household starts life at age 1 and lives 50 periods, after which death is certain[13]. Therefore the population consists of 50 overlapping generations and the index j denotes age. Agents have a probability to survive in each period denoted as s_j and decreasing in age. Different specifications for s_j, corresponding to alternative assumptions regarding annuity markets, are considered. When annuity markets are absent we use a random bequest mechanism to redistribute left-over assets: with negative borrowing limits this opens up the possibility that people dye in debt[14]. Agents are faced with education choices and base such choices on returns and costs of education given age, asset holdings, permanent characteristics and labor shocks. Over their life cycle they choose the labor supply path that maximizes their expected lifetime utility. Education groups are denoted by $e \in \{e_1, e_2, e_3\}$, with $e = e_1$ the lowest and $e = e_3$ the highest. We denote individual permanent characteristics by $\theta \in [\theta_{\min}, \theta_{\max}]$ and let $\{z\}_{j=1}^{50}$ be a sequence of uninsurable idiosyncratic efficiency shocks.

In each period, agents choose their labor supply $n_j \in [0, 1]$ and consume remaining leisure $l_j = 1 - n_j$. They also choose how much of their income to carry over to next period in the form of assets a_{j+1}. We assume that the amount of leisure enjoyed by students is not a choice but rather a deterministic (increasing) function of their permanent characteristics, defined as $l = f^S(\theta)$. Assets bear a risk free net return equal to r, there is a borrowing constraint such that $a_j \geq a_{\min}$ for every j and a transversality condition for agents reaching maximum age such that $a_j \geq 0$ if $j > 50$. Aggregate physical capital is denoted as K and depreciates at rate δ.

Period utility $u(c, l)$ is concave in consumption and leisure; it satisfies standard regularity conditions and in particular the Inada conditions[15]. Future utility is discounted

[13] No explicit retirement is considered, although people can choose not to supply any labor at later ages. We experiment with different labor life lenghts.

[14] Lending without collateral to a person with a mortality risk is equivalent to providing insurance, as argued by Yaari (1965).

[15] In the simulation we use utility of the CRRA class and are of the following type

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

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

by a factor $\beta > 0$. The efficiency weighted labor supply of an agent is defined as

$$h_j = \epsilon_j(\theta, e, z) n_j = \exp\left(\theta + \xi_j^e + z_j\right) n_j \tag{1}$$

where ξ_j^e is an efficiency profile depending on age and education.

The wage rate per efficiency unit of labor in education group e is denoted as w_e. Agents pay proportional taxes τ_n and τ_k on, respectively, labor and asset income: the distinction is kept to separately identify the effect of capital taxation on HC accumulation, which summarises the substitutability between investments in education as opposed to assets[16].

Schooling has a direct cost D_e and is subsidised through a transfer T_e. Individual bequests received by an household at age j as q_j.

When annuity markets are absent, accidental bequests of assets are redistributed among the youngest according to the density prevailing among those who died; alternative initial conditions for wealth are also considered[17]. When annuity markets are absent, an exogenous initial wealth distribution is imposed.

The law of motion for the labor efficiency shocks is summarized by a transition function π denoted as $\pi_{z_{j+1|}z_j} = \pi\{z_{j+1} \mid z_j\}$.

Given some initial conditions $\bar{x}_1$ for the state variables, the age 1 household'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

$$U(\bar{x}_1, c, l) = E_{z \in Z} \sum_{j=1}^{50} S_j \beta^{j-1} u(c_j, l_j) \tag{2}$$

where $S_j = \left(\prod_{i=1}^{j} s_i\right)$. The period budget constraint is

$$c_j + a_{j+1} = Ra_j + \tilde{w}_e n_j (1 - \mathrm{d}_j) - (D_e - T_e)\, \mathrm{d}_j{}^{18} \tag{3}$$

where $R = [1 + r(1 - \tau_k)]$, $\tilde{w}_e = w_e \exp^{\epsilon_j}(1 - \tau_{n^e})$ and d_j is a binary variable which is 1 if the agent is in education and 0 otherwise.

2.2 Household's problem

Education choices depend on relative life-cycle returns to different education levels, current pecuniary cost of schooling, utility cost of studying, current labor shock and asset holdings (z and a)[19].

[16]This effect was first noted by Heckman (1976). If we think of investment as a way to transfer resources intertemporally, changing the price of intertemporal substitution affects the quantity and quality of investments.

[17]Gale and Scholtz (1994) show that inter vivos transfer for education represent only a part of total bequest. We ignore this issue in this paper and redistribute all left over assets among the youngest.

[19]To keep the model as simple as possible we do not explicitly model the sector producing education. Instead we assume that all payments made by agents towards their education are implicitly transformed into improvements in their education status.

To pass from education level e_1 to education level e_2 an agent has to spend in school j_{e_2} successive periods[20], and to pass from e_2 to e_3 an agent has to stay continually in school for j_{e_3} periods[21]. No schooling is possible after e_3 has been achieved[22].

Given prices and direct cost of schooling, the binary function $\mathrm{d}_j = \mathrm{d}_j(\theta, e, z, a)$ describes schooling choice as a mapping from the space of individual states into the age j employment set $\{0, 1\}$.

Individual bequests are also an individual state variable, but we model them as initial (start-of-life) assets in order to reduce complexity. 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 intratemporal margin condition it is possible to express the individual labor supply as a function of optimal consumption and real wage[23].

2.3 Optimality and value functions

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.

We use value functions to characterize the optimal path[24]. A functional equation is an equivalent and unique approach to the household's sequence problem.

The functional equation can be written as

$$J(x_j, p_{j-1}) = \sup_{p_j} \mathrm{v}(p_j) + s_j \beta^{j-1} \int_Z \pi_{z_{j+1}|z_j} J(x_{j+1}, p_j)\, dz_{j+1} \tag{4}$$

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

Gallipoli (2004) shows that a a value function $J^*(x_j, p_{j-1})$ satisfying the functional equation (4) exists and we call $J^*(x_j, p_{j-1})$ the unconditional value function because it is defined over all the possible education choices.

In order to fully characterize the unconditional value function it is helpful to consider 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 $\mathrm{d}_j = 1$.

[20] An additional state variable in the dynamic optimization problem of each agent is therefore the number of previous years of education already under their belt: for simplicity we do not use additional notation for it, although this variable implicitly determines future budget constraints and utility.

[21] In the calibration exercise we set the lenghts of the required study spells to match the main features of the educational system under investigation.

[22] We restrict agents not to go back to school after becoming employed, but we also run some experiments where this restriction is removed.

[23] The analytical details of the labour/leisure intertemporal choice are provided in the 'Preferences' section of the Appendix.

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

We denote the *conditional* value function as $J^*(x_j, p_{j-1} \mid \textit{condition})$, with the condition being the value of d_j[25]. The unconditional functional equation $J^*(x_j, p_{j-1})$ is the upper envelope of the conditional values of employment and education. Without loss of generality, we can reduce the complexity of the value functions associated to different kinds of employment by making the choice of employment sector irreversible (this is equivalent to assume that the costs of reverting to different, feasible 'careers' are sufficiently high). Of course if agents have to possibility to return to education after a working spell they can choose a different employment sector.

The conditional value of employment is denoted as $J^*(x_j, p_{j-1} \mid \mathrm{d}_j = 0) = W_j(\theta, e, z, a)$[26] and is unique. If we restrict agents to never return to education after working spells, this value is defined as

$$W_j(\theta, e, z, a) = \max_{a', n} \; u(c, 1-n) + s_j \beta \int_Z \pi_{z'|z} \, W_{j+1}(\theta, e, z', a') \, dz' \tag{5}$$

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

$$J^*(x_j, p_{j-1} = (1, a_j) \mid \mathrm{d}_j = 0) = \max_e \{W_j(\theta, e, z, a)\}_{e=e_1}^{e_{\max}} \tag{6}$$

where $e_{\max}$ is the agent's 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 (5).

In this case 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.

The next step is to examine the conditional value of education, that we call V_j. The conditional value $V_j(\theta, e, z, a)$ for education participants with $e < e_3$ exists, is unique and is defined as[27]

$$\begin{aligned} J^*(x_j, p_{j-1} \mid \mathrm{d}_j = 1) = V_j(\theta, e, z, a) = \max_{a'} \; u(c, f_e(\theta)) + \\ + s_j \beta \int_Z \pi_{z'|z} \; \max\left\{V_{j+1}(\theta, e', z', a'), \{W_{j+1}(\theta, e, z', a')\}_{e=e_1}^{e_{max}}\right\} dz' \end{aligned} \tag{7}$$

where e_{max} is the education level of the agent. Both V_j and W_j are subject to (3). The conditional value of education for people with $e = e_2$ and in their last year of education is such that $V_{j+1}(\theta, e', z', a') < W_{j+1}(\theta, e, z', a')$ for any e, which satisfies the assumption that no further schooling is possible for people with the highest level of education.[28]

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

[26]The full expression of this value function is $W_j(\theta, e, z, a, ; \mu, R)$ because the distribution of the population over states μ and the exogenous interest rate R , that are given at the beginning of each time period, determine the returns to different production factors.

[27]Also in this case the extensive form of the value expression is $V_j(\theta, s, z, a; \mu, R)$.

[28]Some interesting properties are associated to the value function of education. Details in Appendix.

The education value is not generally concave, but only piece-wise concave with respect to assets. This is potentially troubling because it raises issues of non-uniqueness of the optimal policies. This and related issues are discussed in the Appendix.

When agents are allowed to go back to education, the discounted value of the future in (5) must be extended to include the value of additional education.

Therefore, the unconditional choice problem of an agent with $e = e_2$ is

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

We call this the unconditional problem because we are not restricting the value of d_j. Given a set of individual states, it might happen that there exist two different $\{a_{j+1}, d_j\}$ such that $V_j = W_j$. The corresponding state-space locations are special types of switch points of the unconditional choice problem: they occur if one conditional value function strictly dominates the other everywhere but at the switch points,where they are equal. This sgives the unconditional value function a peculiar butterfly shape.

To guarantee that education decisions are always uniquely determined, we assume that whenever the present value of education is at least as large as the present value of employment, education is chosen over employment. Using this assumption and the set of results obtained for the conditional optimal policies, we argue that the unconditional optimal policy is uniquely determined and piecewise continuous.

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.

2.4 Aggregate variables

We study equilibrium allocations and assume a stationary population. We summarize the aggregate state of the economy by looking at aggregate physical capital K, efficiency-weighted labor supplies (referred to as human capital aggregates) H_1, H_2, and H_3, and defining the measure space $(X, \mathcal{F}(X), \psi_j)$, where X is the individual state space.[29]

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

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

which is a measure of agents belonging to age group j with individual state vector $(\theta, e, z, a) \in F$.[30]

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

[30] 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 $B \in \mathcal{F}(\Delta)$, $\mu(B)$ indicates the mass of agents whose individual state vectors lie in B.

The aggregate states determine the marginal relative return to physical capital and human capital varieties in the economy.

We also assume that the distributions of permanent individual characteristics θ and of idiosyncratic shocks z are independent of time and cohort. 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 normalised to sum up to 1 and are such that $\zeta_{j+1} = s_j \zeta_j$.

2.5 Markets structure

We use the unique good as the numeraire. Such good can be either consumed or saved. In this economy savings a represent ownership rights over physical capital K. We do not model entrepreneurial choices directly, but we maintain that entrepreneurs behave optimally in managing firms. We let the interest rate r be exogenous so that total asset holdings in the economy $\sum_j \zeta_j \int_{\bar{A}} a_j \; d\psi_j(a)$ does not necessarily equal the amount of physical capital K. We residually define the difference between total asset holding and physical capital employed as $FX(r) = K(r) - \sum_j \zeta_j \int_{\bar{A}} a_j \; d\psi_j(a)$.

Positive FX represents the amount of foreign-owned capital that is present in the economy, whereas negative FX represents the amount of foreign activities owned by households.

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.

Given differential mortality and life-cycle assets savings, the various age groups bequeath different assets distributions to the new borns, so that

$$\psi_1(a) = \sum_{j=1}^{\bar{j}-1} \frac{\zeta_j (1 - s_j)}{\sum_{j=1}^{\bar{j}-1} \zeta_j (1 - s_j)} \psi_{j+1}(a) \tag{8}$$

where $\psi_j(a)$ denotes the age j marginal assets density[31][32].

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

$$\begin{aligned} E(q_1) &= \int_{\bar{A}} \psi_1(a) \, a_1 da \\ q_j &= 0 \quad \text{for} \quad j = 2, ..., \bar{j} \end{aligned} \tag{9}$$

[31] This is defined as

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

[32] This bequest mechanism has the desirable feature of making the age 1 assets density depend on the older ages assets densities generated in equilibrium, but we also experiment with exogenous initial assets densities.

and the amount of wealth that is bequetted in each period is[33]

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

In the simulations we also experiment with different degrees of correlation between start-of-life asset holdings and idiosyncratic talent θ, and consider cases in which $\psi_1(a,\theta) \neq \psi_1(a)\,\psi_1(\theta)$. Imposing different patterns of dependence between such marginal densities turns out to be useful if ability is correlated with socio-economic background factors such as family wealth.

We do not consider involuntary unemployment in this model, however people can choose to consume all their leisure endowment if, for example, the market value of their time is too low.

2.6 Technology

Firms maximize profits using a constant returns to scale technology and set wages competitively. The inputs to the aggregate production function are physical capital and three kinds of HC corresponding to different levels of schooling, so that $F(H, K)$ with $H = \{H_1, H_2, H_3\}$. The relationship between human and physical capital is expressed as a Cobb-Douglas function with a nested component modelling the interaction among different kinds of HC:

$$F(H, K) = \bar{A} H^{1-\alpha} K^{\alpha} \tag{10}$$

$\bar{A}$ is an aggregate productivity coefficient[34] and the general, unconstrained definition of the HC input is

$$H = \{A_1 H_1^{\rho} + A_2 H_2^{\rho} + A_3 H_3^{\rho}\}^{\frac{h}{\rho}} \tag{11}$$

with $h = 1$ given the CRS assumption.[35]

In this specification (A_1, A_2, A_3) are share parameters, while ρ pins down the Allen elasticity of substitution among different weigthed labor inputs. In the CES case, we the Allen elasticity of substitution between any two inputs is $\frac{1}{1-\rho}$.[36] When ρ is equal to zero the technology is Cobb-Douglas, whereas values of ρ greater than zero indicate more substitutability than in the Cobb-Douglas case.

[33] In reality, only a part of intra-family wealth transfers are intra-vivos. For a discussion of related issues see Gale & Scholtz, 1994.

[34] In the simulations we normalize $\overline{A}$ to one.

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

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

An alternative and interesting specification is[37]

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

which 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.

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 $r + \delta = \frac{\partial F}{\partial K}$. The total stock of human capital of type e, H_e, is the sum of the efficiency weighted individual labor supplies of type e

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

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

We can also express the marginal condition for physical capital $F_K = r + \delta$ as

$$r + \delta = \frac{\partial F}{\partial K} = \alpha \left(\frac{H}{K} \right)^{1-\alpha} \tag{12}$$

2.7 Government

We assume that the government obtains its revenues from proportional taxation of labor and asset income at respectively τ_{n^e} and τ_k rate, and uses part of the revenues to subsidise education via a transfer T_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

$$\begin{aligned} G + \sum_j \zeta_j \int_X T_e \mathrm{d}_j(x)\; d\psi_j(x) = & \quad (13) \\ = \sum_j \zeta_j \int_X \left[1 - \mathrm{d}_j(x)\right] \tau_{n^e} w_e h_j(x)\, d\psi_j(x) + \sum_j \zeta_j \int_{\bar{A}} r \tau_k a_j\, d\psi_j(a) & \end{aligned}$$

which requires that expenditures equal revenues obtained from taxation[38].

3 Equilibrium

We use a notion of equilibrium in which the measure $\mu(x, j)$ remains unchanged over time. This notion of equilibrium is known as stationary recursive competitive equilibrium

[37] Hamermesh (1993) attributes this 'grouping' production function to Sato (1967).

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

($SRCE$, Lucas, 1980). In Appendix A 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 $x \in X$.

3.0.1 Equilibrium definition

Given an exogenous interest rate r, equilibrium definitions in the asset and good markets must include cross border asset holding FX. Let $(X, \mathcal{F}(X), \psi_j)$ be an age-specific measure space with state space $X = \Theta \times \Im \times Z \times \bar{A}$ and $\mathcal{F}(X)$ a σ-$a\lg ebra$ on X.

Given some state vector $x \in X$ and r, a stationary 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 $V_j(x)$, $W_j(x)$, price functions $w_e(\mu, r)$, $e \in \Im$, densities $(\psi_1, ..., \psi_{\bar{j}})$ and $(\zeta_1, ..., \zeta_{\bar{j}})$, 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, given r.
2. $V_j(\theta, e, z, a)$, $W_j(\theta, e, z, a)$ are the associated value functions.
3. Firms choose capital and human capital in such a way that

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

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 $\mathrm{d}_j(x)$, $a_{j+1}(x)$, $c_j(x)$. Given ζ_j , also $\mu(x,j) = \zeta_j \psi_j(x)$ is a stationary measure.
5. The following equalities hold in the good, asset and labour markets

$$G + \sum_{j=1}^{\bar{j}} \zeta_j \left[\int_X c_j(x)\, d\psi_j(x) + \int_X a_{j+1}(x)\, d\psi_j(x) \right] + rFX =$$
$$= F(H,K) + \sum_j \zeta_j \int_{\bar{A}} a_j\; d\psi_j(a) - \delta K \; - \sum_j \zeta_j \int_X \mathrm{d}_j(x)\, D_e\; d\psi_j(x)$$

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

$$\sum_j \zeta_j \int_X \exp^{\epsilon_j(\theta,e,z)} n_j(x)\, d\psi_j(x) = H_e \qquad \forall e \in \Im$$

We have derived the goods market clearing equation by integrating the individual budget constraint; in the Appendix we show the analytical steps that deliver the goods market equilibrium condition.

4 Identification and Estimation

We assign values to a set of production and efficiency parameters by directly estimating them from US data. This section describes the procedures used to identify and estimate:

- education specific age-earning profiles and skill prices;
- idiosyncratic labor shocks and their law of motion;
- the empirical density of idiosyncratic permanent ability θ over the working population;
- the aggregate technology parameters determining shares and substitution elasticities for different aggregate inputs.

Different data sets are used in the process. It must be stressed that we refer to ability as a set of both observable and unobservable characteristics that have a direct impact on households' earnings but are not explicitly modelled.

4.1 Estimating wage equations: skill prices and age profiles

Skill prices and age-earning profiles can be estimated by imposing some structure upon the data, which we do by using our equilibrium model. For each education group we study a wage equation that is consistent with the individual earning mechanism in our model. The (log-linear) specification of individual hourly wages is therefore

$$\ln w_{it} = w_t + g\left(age_{it}\right) + u_{it} \tag{14}$$

where $u_{it} = \theta_i + z_{it} + m_{it}$. In this notation w_{it} denotes the observed hourly wage rate for individual i at time t, w_t is a (time dependent) hourly return to the specific human capital type, θ_i is an individual ability component, $g\left(age_{it}\right)$ is some function of age and z_{it} is an idiosyncratic transitory shock, possibly autocorrelated. Finally, the term m_{it} denotes measurement error components in wage rates, which for identification purposes are assumed to be uncorrelated across time and orthogonal to all observed and unobserved characteristics.

Our steady state model does not include any time variation, however time varying prices w_t are of pivotal importance for the identification of age effects, residual terms and, in a different context, to pin down the time series of HC aggregates. Since we are primarily interested in non-demographic determinants of educational decisions, we do not model cohort effects explicitly. Furthermore, idiosyncratic abilities are correlated with both educational choices and observed wage rates. By estimating a distinct wage equation for each education group we control for the education self-selection problem, but heterogeneity in unobserved characteristics still represents an obstacle to identification[39].

Assuming linearity of permanent error components, we identify our model parameters by adopting a within group specification for wage equations. We estimate the following specification

$$(\ln w_{it} - \ln \overline{w}_i) = (w_t - \overline{w}) + g\left(age_{it} - \overline{age}_i\right) + (u_{it} - \overline{u}_i) \tag{15}$$

where the bars denote (individual) time averages. This delivers unbiased and consistent estimates of time effects and age profiles[40].

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. All interviews are retrospective, providing data on the previous year.

The sample for this study combined a cross-section sample of nearly 3,000 families, representative of the US population, selected from the Survey Research Center's master sampling frame, and a subsample of about 1,900 families interviewed previously by the Bureau of the Census for the Office of Economic Opportunity. The subsample drawn from the OEO-Census study was limited to low-income families, and compensatory weights were developed in 1968 to account for the different sampling rates used to select the OEO sample component as opposed to the SRC component[41]. A subsample of Latino (Latin American origins) families was added in 1990 and dropped in 1995. Additional immigrant families were added in 1997 and 1999. Moreover, in 1997 some families belonging to the OEO-Census sample component were dropped.

We do not use individuals associated with the Census low income sample, the Latino sample or the New Immigrant sample[42] and focus instead on the SRC core sample, which did not suffer any substantial additions or reductions between 1968 and 2001 and was originally representative of the US population.

[39] We do not include return to experience. Experience is the difference between age and years of schooling, and agents belonging to a given education group have roughly the same number of years of schooling. Therefore the age effects end up capturing returns from experience as well as seniority.

[40] A normalization of the estimation results is necessary to obtain age-earning profiles, skill prices and estimates of permanent heterogeneity. These are then used in numerical simulations. The normalization is bound to be arbitrary because we don't have any 'metric' to measure and compare the relative contributions of age, skills and permanent heterogeneity in determining the final wage rate. A description is included in the Appendix.

[41] In fact the original 1968 wave data must be weighted unless one uses only the SRC representative cross section sample.

[42] Lillard and Willis (1978) make the case that the SEO low income sample should be dropped because of endogenous selection problems.

The main earnings' variable in the PSID refers to the head of the household[43] and is described as total labor income of the head[44]. We use this measure, deflated by the CPI-U for all urban consumers, as the reference earning variable. By selecting only heads of household we ignore other potential earners in a family unit and restrict our attention to people with strong attachment to the labor force[45]. Using heads to approximate households' behavious finds some support in recent work by Hyslop (2001), who provides evidence of very strong and positive assortive matching by couples[46] and shows that such matching is based on permanent individual characteristics. Moreover, the permanent idiosyncratic characteristics upon which the assortive matching depends show significant and strong positive correlation with family permanent income and earnings. We conclude that the sample of households' heads provides a good approximation to the real households' distribution over permanent characteristics, earnings and income.

Information on the highest grade completed is used to allocate individuals to three education groups: less than high school, high school graduates and college graduates, respectively denoted as LTHS, HSG and 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.

Based on our final PSID sample, figures () and () plot the evolution of mean and variance of log annual earnings over time for each education group (LTHS, HSG and CG) and for the whole sample (ALL). The plots show that average annual earnings have experienced a drop in real values during the early 1980s. A steep recovery occurred during all of the 1990s, but this was not sufficient to bring real earnings of lower education groups back to their pre-1980s levels[47]. The 1980s' drop in real earnings was accompanied by higher dispersion: a visual inspection suggests, however, that the variance of earnings of the least educated has not substantially changed in the 30 years span considered, whereas for college graduates there seems to be some upward trend in earnings dispersion. As college graduates represent an ever increasing share of the workers over time, the pooled variance of earnings seems to be upward trended. This behaviour of the pooled variance is consistent with previous empirical evidence.

[43] 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.

[44] 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.

[45] In this way we exclude those individuals who tend to participate only during expansions and whose transitory wage component tends to be relatively larger.

[46] Positive assortive matching means that individuals tend to find partners who have similar permanent characteristics such as education, taste for leisure and cognitive ability. This evidence is based on a sample of US couples observed in the PSID over the 1979-1985 period.

[47] This can be in part explained by progressive self-selection of more able people out of the lowest education group.

(Figure 1)

We use real hourly wage rates as dependent variable in the wage equations. Figures () and () report the evolution of hourly wage rates for the sample groups. While the pattern of the average hourly wage rates over time is very similar to the one of yearly earnings, the evolution of variance over time is much more trended, especially for the group of High School graduates. This can be interpreted as evidence that labor supply plays some role in smoothing earnings over time.

(figure 2)

(figure 3)

We estimate a wage specification as in(15) and use a 4^{th} degree polynomial in age to approximate the possibly non linear $g(age_{it})$ functions. The age polynomials' coefficients are presented in table (1).

Table 1: Age polynomials' coefficients

Dependent variable: log hourly earnings		
coeff.	point estimate	S.E.
	Education=LTHS	
age	0.0412505	.0081143
age^2	-0.0004179	.0000905
	Education=HSG	
age	0.4928285	0.1071015
age^2	-0.0162768	0.0039883
age^3	0.0002413	0.0000644
age^4	-1.34e-06	3.82e-07
	Education=CG	
age	0.8697329	0.1560285
age^2	-0.0282	0.0058548
age^3	0.0004149	0.0000953
age^4	-2.30e-06	5.69e-07

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

(figure 4)

By fitting the within group specification of the wage equation we also obtain estimates of the year specific price effects, which are plotted in figure (). The time effects have a natural interpretation as time varying prices of skills associated to different education groups and can be used to identify the supply of human capital in the economy. In the Appendix we describe the procedure that we use to normalize the estimates of age, time and fixed effects.

(figure 5)

4.3 HC Aggregates and Education Specific Labor Inputs

Consider the problem of studying the production function in equation (10). From national accounts data we can obtain long time series of aggregate output

$$Y_t = F(H_{1t}, H_{2t}, H_{3t}, K_t)$$

We also have observations on aggregate physical capital, K_t, and the wage bills that are paid in each year to different education groups (denoted as WB_t^e)[48].

In order to retrieve technology parameters it is necessary to identify and estimate HC aggregates, H_{et}[49], which are defined (see eqt. **??**) as an efficiency-weighted sum of individual labor supplies. The crucial question is whether we can recover (H_1, H_2, H_3) for a reasonably long number of periods.

Our approach is to identify (H_1, H_2, H_3) by combining the observable wage bills (WB^1, WB^2, WB^3) and the estimated year specific skill prices, w_{et}. In fact, by definition we have that

$$WB_t^e = w_{et} H_{et} \tag{16}$$

and we can identify HC aggregates by using the estimated time series of skill prices $\widehat{w_{et}}$ obtained from the wage equations.

The main problem with this identification procedure is a data measurement problem: in the US fringe benefits and other employer's contributions are often not recorded as straightforward earnings[50] and can account for a sizeable proportion of yearly earnings. Furthermore, they are likely to represent different proportions of total earnings in different education groups, and tend to be higher for college graduates. This might lead to and underestimate of the aggregate human capital for the higher education groups.

4.4 Combining CPS data and PSID estimates

(figure 6)

The wage bills should reliably represent the distribution of US working population over education groups in each year. For this reason we use 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.[51] 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

[48] The (yearly) wage bill for a given education group is the total earning payments received by people of that education group in a given year.

[49] The subscripts e and t stand for human capital type (as implied by the educational level) and date of observation.

[50] An example, pointed out to us by Ken Judd, is that of employer's pension contributions which can account for over 10% of yearly earnings.

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

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 noninterview 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. 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. 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.

We use the CPI for all urban consumer (with base year in 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, we have extrapolated the average of the top-coded values by using a tail approximations based on a Pareto distribution. This procedure is based on a general approach to inference about the tail of a distribution originally developed by Bruce 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 (2002) provides evidence that this method closely approximates the average of the top-coded tails by using undisclosed and confidential non top-coded data available at the BLS.

Figure () reports the number of people working in each year by education group. It is clear that some strong and persistent trends towards higher levels of education have characterized the sample period. Figure () plots both the average earnings by year and education group in levels based on the CPS. 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,1984) which is based on the original Hill's estimator to approximate exponential tails. The difference between the two averages is larger for the most educated people who tend to be more affected by top-coding. We include also self-employed people in the computation of these aggregates; however, their exclusion has almost no effect on the value of the wage bills and human capital aggregate, as they never represent more than 5% of the working population in a given education group (and most of the times much less than that). Figure () plots the wage bills (in billion of 1992$) by year and education group. Dividing the wage bills by the exponentiated value of the time effects estimated through the wage equations using PSID data we finally obtain the human capital aggregates, that are plotted in figure ().

(figure 7)

Using the time effects $\widehat{w_{et}}$ estimated through the wage equations we provide point estimates of the different HC aggregates by year: this estimates are presented in figure ().

(figure 8)

4.5 Distribution of permanent characteristics

We 'residually' identify ability θ_i, whose estimate is denoted as $\widehat{\theta}_i$, from the sequence of agent-specific residuals associated to the wage equation (14)[52]. We resort to the fact that

$$\ln w_{it} = w_t + \theta_i + g\left(age_{it}\right) + e_{it}$$

and compute

$$\widehat{\theta}_i = \frac{\sum_{t=1}^{T(i)} \ln w_{it} - \widehat{w}_t - \widehat{g\left(age_{it}\right)}}{T\left(i\right)} \approx \theta_i$$

where $T\left(i\right)$ is the total number of observation 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 distribution over the working population of the ability to earn.

Under this specification the individual fixed effects θ_i capture all omitted sources of permanent heterogeneity which have some effect on individual earnings: they range

[52]The so-called incidental parameters problem affects the FE estimates: limited panel lenght is responsible for FE estimation bias. The fact that our estimates rely on people who are observed at least for 8 years will partially reduce the bias.

from observable characteristics such as gender, cohort and race to non-observable characteristics such as cognitive ability and family incentives; in this sense, the resulting distribution of estimated fixed effects can be thought of as a single-index summary of multi-dimensional heterogeneity.

These forms of heterogeneity constitute an essential part of the individual ability to earn that is not due to age or price effects and it is important to include such forms of heterogeneity in the estimated measure of idiosyncratic ability which are used in numerical simulations[53].

We have estimated this distribution using both the large sample covering the period between 1967 and 2000 and the smaller sample for which Release II data are available covering the period 1968-1993. Furthermore, we have checked whether weighting the estimated $\widehat{\theta}$ using PSID longitudinal weights would change the estimated empirical frequencies of ability.

(figure 9)

In figure () we report the empirical frequencies of $\widehat{\theta}$. Changing the lenght of the sample and using weights does not introduce any substantial variation on both shape and location of the density.

4.6 Analysis of labor efficiency shocks

The wage equation residual, rescaled by removing the permanent ability component, varies along time and across individuals and is defined as

$$\widetilde{u}_{it} = w_{it} - g\left(age_{it}^{norm}\right) - w_t^{norm} - \theta_i^{norm}$$

We assume that $\widetilde{u}_{it}$ can be decomposed as

$$\widetilde{u}_{it} = z_{it} + m_{it}$$

where z_{it} is an autocorrelated error process and m_{it} is classical measurement error $iid\left(0, \sigma_m^2\right)$.

If we model the autocorrelated z process as

$$z_{it} = \rho z_{it-1} + \varepsilon_{it}$$

with $\varepsilon_{it}\tilde{}\, iid\left(0, \sigma_\varepsilon^2\right)$, we can achieve identification of the autoregressive parameters in one of several ways. A first possibility is to use the following second moments

[53] It must be noticed that our estimates can provide an estimate of the unconditional distribution of the ability to earn, but cannot be used to infer information about the conditional distributions of ability in each education group because the longitudinal PSID sample that we select fails to represent the marginal densities of the US population over age and education. This problem becomes more severe the further away we go from the original sampling date. However, granted that our sample selection technique is mostly orthogonal with respect to ability, we can confidently use the estimated emnpirical density as an approximation.

$$VAR(\widetilde{u}_{it}) = VAR(z_{it}) + VAR(m_{it})$$
$$COV(\widetilde{u}_{it}, \widetilde{u}_{it-1}) = COV(z_{it}, z_{it-1})$$

where

$$VAR(z_{it}) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$
$$COV(z_{it}, z_{it-1}) = \frac{\rho\sigma_\varepsilon^2}{1-\rho^2}$$

and compute[54]

$$\rho = \frac{COV(z_{it}, z_{it-1})}{VAR(z_{it})} = \frac{COV(\widetilde{u}_{it}, \widetilde{u}_{it-1})}{VAR(\widetilde{u}_{it}) - VAR(m_{it})} \quad (17)$$

Of course an external estimate of $VAR(m_{it})$ is necessary in this case. An alternative way to identify the autoregressive coefficient without resorting to an external estimate of $VAR(m_{it})$ is available whenever we can assume classical measurement error: in fact, this implies

$$\rho = \frac{COV(z_{it}, z_{it-2})}{COV(z_{it}, z_{it-1})}$$

We can estimate both a unique autoregressive coefficient ρ for all education groups and a set of group specific ρ_{edu}.[55]

Furthermore, we can compute

$$\widetilde{u}_{it} - \hat{\rho}\widetilde{u}_{it-1} = \hat{\varepsilon}_{it} = (z_{it} - \hat{\rho}z_{it-1}) + (m_{it} - \hat{\rho}m_{it-1}) = \varepsilon_{it} + (m_{it} - \hat{\rho}m_{it-1}) \quad (18)$$

The moments of the constructed residual $\hat{\varepsilon}_{it}$ are

$$VAR(\hat{\varepsilon}_{it}) = \sigma_\varepsilon^2 + \left(1+\hat{\rho}^2\right)\sigma_m^2$$
$$COV_1(\hat{\varepsilon}_{it}) = -\hat{\rho}^2\sigma_m^2$$
$$COV_j(\hat{\varepsilon}_{it}) = 0 \quad j \geq 2$$

and can be used to test the goodness of the specification we assume for the z process.

[54] In fact if the above specification is correct then we have that

$$VAR(\widetilde{u}_{it}) = \frac{\sigma_\varepsilon^2}{1-\rho^2} + \sigma_m^2$$

$$COV_j(\widetilde{u}_{it}) = \frac{\rho^j\sigma_\varepsilon^2}{1-\rho^2} \quad j \geq 1$$

[55] Estimates are based on non weighted residuals, as weighting would not add any information, since heterogeneity is factored out of the errors by construction.

4.7 Estimation and Testing of Labor Shock Processes

The estimated autoregressive (ρ) coefficients and their (bootstrapped) standard error are presented in the following table. Higher persistence is associated with higher values for $\hat{\rho}$: higher persistence can be interpreted as a less insurable kind of shock and corresponds to a more volatile lifecycle pattern for earnings.

Table 2: Estimates of the autoregressive coefficient $\hat{\rho}$, by education group and pooled. 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)

The estimated values for $\hat{\rho}$ seem to indicate that group 2 (High school graduates) experience the lowest earnings' risk. The associated

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

$$\varepsilon_t = \ln F(H_t) - \rho \ln F(H_{t-1}) - X_t'\beta + \rho X_{t-1}'\beta \tag{19}$$

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

Table 3: Autocovariances of labor shocks for Education group 1 (LTHS), by year and pooled. Asymptotic standard error in parenthesis.

year	Lag 0	Lag 1	Lag 2	Lag 3	year	Lag 0	Lag 1	Lag 2	Lag 3
1968	.0604				1983	.0626			
	(.008)					(.0120)			
1969	.0757				1984	.0919			
	(.0137)					(.0199)			
1970	.0490				1985	.0849			
	(.0071)					(.0188)			
1971	.0541				1986	.0983			
	(.0119)					(.0188)			
1972	.0471				1987	.0915			
	(.0109)					(.0167)			
1973	.0402				1988	.0967			
	(.0047)					(.0215)			
1974	.0446				1989	.1148			
	(.0056)					(.0294)			
1975	.0604				1990	.0687			
	(.0084)					(.0121)			
1976	.0696				1991	.0762			
	(.0107)					(.0145)			
1977	.0780				1992	.0844			
	(.0198)					(.0147)			
1978	.0729				1993	.1356			
	(.0121)					(.0473)			
1979	.0609				1994	.0638			
	(.0079)					(.0108)			
1980	.0674				1995	.1016			
	(.0109)					(.0355)			
1981	.0518				1996	.0620			
	(.0087)					(.0121)			
1982	.0597				1997	.0657			
	(.0072)					(.0171)			
POOLED									

We initially include a time polynomial of the form $t\,(time,\gamma) = c + \gamma_1 time_t + \gamma_2 time_t^2 + \gamma_3 time_t^3$ in the conditional mean of $\ln F\,(H_t)$. 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

$$\varepsilon_t = \ln F\,(H_t) - X_t'\beta - \gamma time_t - \rho\left[\ln F\,(H_{t-1}) - X_{t-1}'\beta - \gamma time_{t-1}\right] \qquad (20)$$

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 $H_{t-1-m-g} = [H_{1,t-1-m-g}\ ,\ H_{2,t-1-m-g}\ ,\ H_{3,t-1-m-g}]$. Given this notation it follows that $3\,(M+1)$ 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 (20) 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 weighting 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 Mead (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[56].

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 $30 - 1 - m - g$.

[56]We also run further check with a standard Powell-Newton algorithm using calculus conditions to identify a minimum.

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.

The first table 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.

Dependent Variable Based on Wage Bills, First Step Weighting 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

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

Dependent Variable Based on Wage Bills, First Step Weighting 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

5 Simulations

(to be completed) Each time unit represents a year and the parameters are based on yearly estimates.

5.1 Preferences parameters

The parameters ν and λ of the period utility (2) jointly pin down the intertemporal elasticity of substitution of consumption, that is $\frac{1}{1-\nu(1-\lambda)}$. With $\nu = 0.33$ and $\lambda = 2.00$ we have that such elasticity is roughly 0.75.

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 $\bar{j} = 50$ years, after which, at the real age of 65, death is certain (the retirement age is not considered in this analysis, so that agents die at the end of their working life). The sequence of conditional survival probabilities $\{s\}_{j=1}^{50}$ is based on mortality tables for the US.

The cost of education D_e is expressed as a proportion of equilibrium earnings, whereas the education subsidy T_e.

5.3 Simulation Results

(to be added)

The most obvious 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 following table reports results of such experiments. At the bottom there are results pertaining to a different kind of experiment based on substantial cuts to either earned income tax rate or unearned income tax rate.

	Tuit. $92	Subs. $92	% workers by edu			Month. Salary $92, pretax			0 assets % of pop.	r %
			LTHS	HS	C	LTHS	HS	C		
Benchmark	5826	0	25.2	58.5	16.3	1111	2023	2899	12.2	3.92
50% subs. (PE)	4831	2415	5.3	16.2	78.5	921	1677	2403	5.7	3.92
50% subs. (GE)	5821	2910	25.2	58.4	16.4	1110	2021	2890	12.1	3.90
150% subs. (PE)	4868	7302	5.3	16.4	78.2	928	1690	2422	6.0	3.92
150% subs. (GE)	5840	8761	25.1	58.3	16.5	1121	2027	2889	11.9	3.87
cut K tax 2/3 (PE)	10411	0	36.8	29.9	33.2	1985	3615	5180	0.08	3.92
cut K tax 2/3 (GE)	5334	0	25.7	54.0	18.6	1015	1852	3079	15.9	4.14
cut L tax 2/3 (PE)	4996	0	6.7	20.6	72.7	953	1735	2486	8.9	3.92
cut L tax 2/3 (GE)	6651	0	20.4	52.4	27.2	1164	2310	1875	18.9	5.3

6 Conclusions

To be added.

Preliminary results are available for the UK. Simulations with US parameters to follow shortly.

References

[1] Abowd, John M. and David Card (1989), 'On the covariance structure of earnings and hours change', Econometrica, 57, 411-445

[2] Acemoglu, Daron (1998), 'Why do new technologies complement skills? Directed technical change and wage inequality', Quarterly Journal of Economics, 113, 1055-1090

[3] Acemoglu, Daron (2000), 'Technical change, inequality, and the Labor Market', manuscript

[4] Abraham, Arpad (2001), 'Wage Inequality and Education Policy with Skill-biased Technological Change in OG Setting', manuscript

[5] Becker, Robert A. and John H. Boyd III (1997), 'Capital Theory, Equilibrium Analysis and Recursive Utility'.

[6] Bound, John and George Johnson (1991), 'Wages in the United States during the 1980s and beyond', inM. Kosters,ed.,Workers and their wages (American Enterprise Institute, Washington DC).

[7] Bound, John and George Johnson (1992), 'Changes in the structure of wages in the 1980s an evaluation of alternative explanations', American Economic Review 82371-392.

[8] Conesa, Juan C. and Dirk Krueger (1999), 'Social Security Reform with Heterogeneous Agents', Review of Economic Dynamics, 2757-795

[9] Cooley, Thomas F. and Edward C. Prescott (1995), 'Economic Growth and Business Cycles', in 'Frontiers of Business Cycle Research' edited by Thomas F. Cooley, Princeton University Press

[10] Costa Dias, Monica (2002), 'A GE approach to the evaluation of social programmes', PhD dissertation, University College London.

[11] Diaz-Jimenez, Javier, Vincenzo Quadrini, Jose' Victor Rios-Rull (1997), Federal Reserve Bank of Minneapolis Quarterly Review, Vol.21, No.2,pp.3-21

[12] Eaton, Jonathan and Harvey S. Rosen (1980), 'Taxation, Human Capital and Uncertainty', American Economic Review, 70705-715

[13] Florens, Jean-Paul, James Heckman, Costas Meghir and Edward Vytlacil (2002), 'Instrumental Variables, Local Instrumental Variables and Control Functions', Cemmap WP19/02

[14] Galor Oded and Joseph Zeira (1993), 'Income Distribution and Macroeconomics', Review of Economic Studies, 6035-52

[15] Galor, Oded and Daniel Tsiddon (1997), 'The Distribution of Human Capital and Economic Growth', Journal of Economic Growth, 293-124

[16] Gosling, Amanda, Sthephen Machin and Costas Meghir (2000), 'The changing distribution of male wages in the UK', Review of Economic Studies 67635-66.

[17] Hansen, Gary D. (1993), 'The Cyclical and Secular Behaviour of the Labour Input-Comparing Efficiency Units and Hours Worked', Journal of Applied Econometrics, 871-80

[18] Heathcote, Jonathan, Kjetil Srorestletten and Giovanni L. Violante (2000), 'From Wage to Consumption Inequality: a Quantitative Analysis', mimeo

[19] Heckman, James J., Lance Lochner and Cristopher Taber (1998), 'Explaining rising wage inequalityexplorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents', NBER working paper no.6384.

[20] Heckman, James 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' edited by Charles F. Manski and Daniel L. McFadden, Cambridge, MIT Press

[21] Carneiro, Pedro and James J. Heckman(2002), 'Human Capital Policy', mimeo, University of Chicago

[22] Heathcote Jonathan, Kjetil Storesletten and Gianluca Violante (2003), 'From Wages to Consumption Inequality: a Quantitative Analysis', mimeo

[23] Hill Bruce M. (1975), 'A Simple General Approach to Approximate the Tail of a Distribution', The Annals of Statistics, Vol.3, No.5, 1163-1174

[24] Huggett Mark and Gustavo Ventura(1997), 'On the Distributional Effects of Social Security Reform', mimeo

[25] Hyslop Dean (2001), 'Rising U.S. Earnings Inequality and Family Labor Supply: the Covariance Structure of Intrafamily Earnings', American Economic Review, Vol.91 No.4, September 2001, p.755-777

[26] Katz, Lawrence F. and David H. Autor (1999), 'Changes in the wage structure and earnings inequality', Handbook of Labor Economics, Volume 3, edited by O. Ashenfelter and D. Card

[27] Kapicka Marek (2002), 'Income Taxation and Human Capital Accumulation', PhD Dissertation, University of Chicago

[28] Katz, Lawrence F. and Kevin M. Murphy (1992), 'Changes in relative wages, 1963-87: supply and demand factors' Quarterly Journal of Economics 10 735-78

[29] Keane, Michael P., Robert Moffitt and David Runkle (1988), 'Real wages over the Business Cycle: Estimating the Impact of Heterogeneity with Micro Data', Journal of Political Economy, vol.96, no.6.

[30] Keane Micheal P. and Kenneth Wolpin (1994), 'The Solution and Estimation of Discrete Choice Dynamic programming Models by Simulation and Interpolation: Monte Carlo Evidence', Review of Economics and Statistics, Vol.76, Issue 4 (Nov.1994), 648-672

[31] Keane Micheal P. and Kenneth Wolpin (1997), 'The career decisions of young men', Journal of Political Economy, vol.105, no.3.

[32] Krussell, Per, Lee E. Ohanian, Jose' Victor Rios-Rull and Giovanni L. Violante (2000), 'Capital-skill complementarity and inequality', Econometrica.

[33] Kyriazidou, Ekaterini (1997), 'Estimation of a Panel Data Sample Selection Model', Econometrica, Vol.65, N.6, pages 1335-1364.

[34] Lee, Donghoon (2001), 'An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice', manuscript

[35] Levhari, David (1974) and Yoram Weiss, 'The effect of risk on the investment in human capital', American Economic Review, Vol.64, pages 950-63.

[36] Meghir, Costas and Edward Whitehouse (1996), Journal of Labor Economics, Vol.14, , no.1

[37] Meghir, Costas and Luigi Pistaferri (2002), 'Income Variance Dynamics and Heterogeneity', Cemmap WP0/07

[38] Olley Steven G. and Ariel Pakes (1996), 'The Dynamics of Productivity in the Telecommunications Equipment Industry', Econometrica, 64 1263-1297

[39] Pavoni, Nicola (2002), 'Optimal Unemployment Insurance with Human Capital Depreciation and Duration Dependence', mimeo, UCL

[40] Polivka, Anne E. (2000), 'Using Earnings Data from the Monthly Current Population Survey', mimeo, Bureau of Labor Statistics

[41] Rudin, Walter (1953), 'Principles of Mathematical Analysis'

[42] Stokey, Nancy L., Robert E. Lucas with Edward Prescott (1989), 'Recursive Methods on Economic Dynamics'

[43] Townsend, Robert M. and Kenichi Ueda (2001), 'Transitional Growth with Increasing Inequality and Financial deepening', mimeo

[44] Trostel, Philip A. (1993), 'The Effect of Taxation on Human Capital', Journal of Political Economy, Vol.101, Issue 2, 327-350

[45] West, Sandra A. (1985), 'Estimation of the Mean from Censored Income Data', Bureau of Labor Statistics Report

[46] Whelan, Karl (2000), 'A Guide to the Use of Chain Aggregated NIPA Data', mimeo, BEA

[47] Wood, Adrian (1995), 'How trade hurt unskilled workers', Journal of Economic perspectives 957-80

[48] Wood, Adrian (1998), 'Globalisation and the rise in labour market inequalities', Unpublished paper(Institute of Development Studies, University of Sussex).

[49] Yannelis, Nicholas C.(1997), 'Lecture Notes in General Equilibrium Theory', manuscript, University of illinois, Urbana-Champaign.

A Definition of Stationary Measure

Stationary measure μ^*

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

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

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

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

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

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

There are conditions which guarantee the existence of a unique fixed point $\psi_j^*(.)$

1. Monotonicity of the decision rules $a(\cdot)$, $e(\cdot)$, $c(\cdot)$; (sufficient condition);
2. Monotone mixing property of the measure μ_i, $\forall i$; (necessary condition).

B Analytical derivation of the market clearing condition

The budget constraint of a generic agent is described in equation (3) as

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

Such expression can be simplified by using equation (9) to express q_j, so that we obtain

$$\begin{aligned} &c_j + a_{j+1} = \\ &= [1 + r(1 - \tau_k)]\, a_j + w_e \exp^{\epsilon_j} n_j (1 - \tau_{n^e})(1 - \mathrm{d}_j) - (D_e - T_e)\,\mathrm{d}_j \end{aligned} \tag{21}$$

where $a_1 = q_1$, with $q_j = 0$ and $a_{j+1} = a_{j+1}(x)$ for $j = 2, ..., \bar{j}$. Notice that $E(q_1)$ is described in (9).

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

$$\sum_{j=1}^{\bar{j}} \zeta_j \left[\int_X c_j(x)\, d\psi_j(x) + \int_X a_{j+1}(x)\, d\psi_j(x) \right] = \tag{22}$$

$$= (1 + r(1-\tau_k)) \sum_{j=1}^{\bar{j}} \zeta_j \int_{\bar{A}} a_j\, d\psi_j(a) +$$

$$+ \sum_{j=1}^{\bar{j}} \zeta_j \int_X w_e \exp^{\epsilon_j} n_j(x)(1-\tau_{n^e})(1-\mathrm{d}_j(x))\, d\psi_j(x) +$$

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

Using the government budget constraint from equation (13) we can rewrite (22) as

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

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

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

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

$$G+\sum_{j=1}^{\bar{j}}\zeta_j\left[\int_X c_j(x)\,d\psi_j(x)+\int_X a_{j+1}(x)\,d\psi_j(x)\right]= \tag{23}$$

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

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

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

$$G+\sum_{j=1}^{\bar{j}}\zeta_j\left[\int_X c_j(x)\,d\psi_j(x)+\int_X a_{j+1}(x)\,d\psi_j(x)\right]= \tag{24}$$

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

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

$$G+\sum_{j=1}^{\bar{j}}\zeta_j\left[\int_X c_j(x)\,d\psi_j(x)+\int_X a_{j+1}(x)\,d\psi_j(x)\right]= \tag{25}$$

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

This is exactly the goods market clearing equilibrium condition.

C FONCs and "Consumption Only" budget sets

Wages

Given the functional form of the transformation function and defining the generic human capital factor H as

$$H=\left\{A_1H_1^{\rho_1}+(1-A_1)\left[A_2H_2^{\rho_2}+(1-A_2)H_3^{\rho_2}\right]^{\frac{\rho_1}{\rho_2}}\right\}^{\frac{1}{\rho_1}}$$

the equilibrium wages can be written in analytical form as

$$w^1=\frac{\partial F}{\partial H_1}=(1-\alpha)\frac{F(H,K)}{H^{\rho_1}}A_1H_1^{\rho_1-1} \tag{26}$$

$$w^2=\frac{\partial F}{\partial H_2}=(1-\alpha)\frac{F(H,K)}{H^{\rho_1}}A_2H_2^{\rho_2-1}M \tag{27}$$

$$w^3 = \frac{\partial F}{\partial H_3} = (1-\alpha)\frac{F(H,K)}{H^{\rho_1}}(1-A_2)\, H_3^{\rho_2 - 1} M \tag{28}$$

where $M = (1-A_1)\left[A_2 H_2^{\rho_2} + (1-A_2) H_3^{\rho_2}\right]^{\frac{\rho_1}{\rho_2}-1}$.

Preferences and Education Choice

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

$$u(c) = \frac{\left[c^{\nu} f^e(\theta)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda} \tag{29}$$

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

The period utility for an employed agent is instead

$$u(c,l) = u(c, 1-n) = \frac{\left[c^{\nu}(1-n)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda} \tag{30}$$

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

$$u_c(c,l) = \left(c^{\nu}(1-n)^{1-\nu}\right)^{-\lambda} \nu c^{\nu-1}(1-n)^{1-\nu} \tag{31}$$
$$u_l(c,l) = \left(c^{\nu} l^{1-\nu}\right)^{-\lambda}(1-\nu)\, c^{\nu} l^{-\nu}$$

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

From the intratemporal 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

$$n = \max\left\{1 - \left(\frac{1-\nu}{\nu}\right)\frac{c}{w_e \exp^{\epsilon^e}(1-\tau_{n^e})}, 0\right\} \tag{32}$$

If we plug equation (32) into the period budget constraint of a working agent,(3), we can cancel out labor supply and obtain a *'consumption only'* budget constraint

$$c_j + a_{j+1} = \left[1 + r(1-\tau_k)\right] a_j + w_e \exp^{\epsilon_j}(1-\tau_{n^e})\left(1 - \frac{1-\nu}{\nu}\frac{c_j}{w_e \exp^{\epsilon_j}(1-\tau_{n^e})}\right)$$
$$\Longrightarrow \nu(c_j + a_{j+1}) = \nu\left[1 + r(1-\tau_k)\right] a_j + \nu w_e \exp^{\epsilon_j}(1-\tau_{n^e}) - (1-\nu)\, c_j$$
$$\Longrightarrow c_j + \nu a_{j+1} = \nu\left[1 + r(1-\tau_k)\right] a_j + \nu w_e \exp^{\epsilon_j}(1-\tau_{n^e}) \tag{33}$$

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

$$\frac{\left[c^{\nu}\left(\frac{1-\nu}{\nu}\frac{c}{w_e \exp^{\epsilon}(1-\tau_{n_e})}\right)^{1-\nu}\right]^{(1-\lambda)}}{1-\lambda} = \frac{\left[\left(\frac{1-\nu}{\nu}\frac{1}{w_e \exp^{\epsilon}(1-\tau_{n_e})}\right)^{1-\nu} c\right]^{(1-\lambda)}}{1-\lambda} \tag{34}$$

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.(33) 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

$$n_j = \max\left\{1 - \left(\frac{1-\nu}{\nu}\right)\frac{\nu R a_j + \nu\widetilde{w} - \nu a_{j+1}}{\widetilde{w}}, 0\right\} \tag{35}$$
$$= \max\left\{\nu + (1-\nu)\frac{a_{j+1} - R a_j}{\widetilde{w}}, 0\right\}$$

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 $(1-\nu)\nu$ 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.

D The 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[57]. 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

[57] 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.

period 1994-2001[58].

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.

D.1 Sample selection

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

The probability sample of individuals defined by the original 1968 sample of PSID families was then followed in subsequent years. A distinction between original sample individuals, including their offspring if born into a responding panel family during the course of the study (i.e., both those born to or adopted by a sample individual), and non-sample 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.

Sample persons who are living members of a 1968 PSID family have a sample selection factor equal to the reciprocal of the selection probability for their 1968 PSID family unit. The computation of the sample selection weight factor for sample persons who are "born into" a PSID family after 1968 uses a formula that is conditional on the "sample status" of their parents. However, data for nonsample persons present a problem for longitudinal analysis since the time series for these individuals is left censored at the date at which they entered the PSID family. Furthermore, it is not likely that this left censoring is random with respect to the types of variables that might be considered in longitudinal analysis. Because of the left censoring of their data series, nonsample persons in PSID families have historically been assigned a zero value selection weight factor and a zero-value for the PSID longitudinal analysis weight[59]. This is of course a problem when using

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

[59] Beginning with the 1993 wave, PSID is providing users with a file that includes special weights that will enable analysts to include all 1993 sample and nonsample person respondents in cross sectional analysis of the 1993 PSID data set. These

the core SRC: non sample people can be tracked through their Person 1968 number (that assumes values between 170 and 228) and whenever we use individual weights we control for the presence of non-sample individuals.

An additional dimension that is included in the core longitudinal weights are adjustments for panel attrition due to nonresponse and mortality. Attrition adjustments were performed in 1969 and every five years thereafter.

In general individual longitudinal weight values for PSID core sample persons are the product of three distinct sets of factors, that can be summarized as follows:

1. a single factor that represents the reciprocal of the probability by which the sample person was "selected" to the PSID panel;
2. a compound product of attrition adjustment factors developed in 1969 and every 5 years thereafter,
3. mortality adjustment factors also developed and applied in 1969 and every 5 years thereafter.

A general formula that reflects the composite nature of the individual weights is:

$$W_{i,1993} = W_{i,sel} \times \prod_{j=1969}^{T} \left[W_{i,NR(j)} \times W_{i,M(j)} \right] \tag{36}$$

where: $W_{i,sel}$ is the selection weight factor – the reciprocal of probability that individual i is selected to the PSID panel by membership in a 1968 PSID sample family or by birth to a PSID sample parent; $W_{i,NR(j)}$ is the attrition adjustment factor applied to the i^{th} individual weight at time period j; $W_{i,M(j)}$ is the age, sex and race-specific mortality adjustment applied to the i^{th} individual weight at time period j[60].

The 1967-1992 Final Release Sample

The 1968-1993 PSID individual file contains records on 53,013 individuals (that is, all who were ever present in the sample at least on one year) We drop members from the Latino sample added in 1990 (10,022 individuals) and keep a sample of 42,991 individuals. We then drop those who are never heads of their household and we are left with a sample of 16,028 individuals. We then drop all individuals who are younger than 25 and older than 60, which leaves us with a sample of 13,399 individuals. Dropping observations for self-employed people reduces the sample to 11,574 individuals.

We keep in our sample only people with at least 8 (possibly non continuous) observations, which leaves us with 4,529 individuals. Dropping individuals with missing, zero or top-coded earnings reduces the sample to 4,300 individuals, and dropping individuals with total hours of work that are missing, zero or larger than 5840 further

weights are called cross-sectional weights (as opposed to the standard longitudinal weights that have been produced from 1969 onwards).

[60] Of course, non sample people have a zero weight because $W_{i,sel} = 0$ for them.

reduces our sample to 4,295 individuals. We eliminate individuals with outlying earning records, defined as changes in log-earnings larger than 4 or less than -2, which leaves 4,211 individuals in the sample.

Finally, dropping people who are connected with the original SEO low-income sample leaves us with a sample of 2,371 individuals.

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

Table 4: Distribution of observations for the 1967-1992 sample, by year

year	**Number of Observations**	**year**	**Number of Observations**
1967	783	1980	1575
1968	853	1981	1551
1969	906	1982	1551
1970	965	1983	1586
1971	1090	1984	1636
1972	1192	1985	1656
1973	1280	1986	1610
1974	1328	1987	1535
1975	1382	1988	1484
1976	1428	1989	1415
1977	1489	1990	1349
1978	1513	1991	1285
1979	1550	1992	1201

Table 5: Distribution of observations for the 1967-1992 sample, by education group

years of education	**Number of Individuals**	**Number of Observations**
less than 12	330	4,804
12 to 15	1,354	19,902
16 or more	687	10,487

Add discussion on sample and non-sample individuals (SEQUENCE # 170-228 - THERE ARE NONE IN OUR SUBSAMPLE !

The 1967-2000 Mixed (Final and Early Release) 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 6: 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 7: Distribution of observations for the 1967-2000 sample, by education group

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

D.2 Testing 2^{nd} Order Moments of the Labor Shocks

After computing autoregressive coefficients for the labor shocks we also provide some tests of the goodness of our specification, based on the covariance matrices of the labor shocks. This paragraph describes the procedure.

Main references: Gary Chamberlain (Panel Data, Chapter 22, Handbook of Econometrics, Volume II, edited by Z.Griliches and M.D.Intriligator, 1984), Abowd and Card (Econometrica,1989), Richard Dickens (Economic Journal, 2000).

Suppose we have observation on some variable X that is indexed by individual and time period, that is we have observations $\{x\}_{it}$, where i denotes an individual between 1 and n, and t denotes a year between 1 and T.

Define a vector

$$d_i = \begin{pmatrix} d_{i1} \\ \vdots \\ \vdots \\ d_{iT} \end{pmatrix}$$

where d_{it} is an indicator variable such that: $d_{it} = 1$ if the individual is present in year t of the panel; $d_{it} = 0$ otherwise; and T is the total lenght of the panel.

Analogously we can define a vector

$$x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ \vdots \\ x_{iT} \end{pmatrix}$$

where x_{i1} can be the error terms we are studying. Since our panel is unbalanced the elements of x_i corresponding to missing years of data are set to zero.

If we define a cell-by-cell scalar product operator between conformable matrices A and B as follows

$$A \dot{\times} B = \begin{pmatrix} A_{11} \times B_{11} & \cdots & A_{1n} \times B_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} \times B_{n1} & \cdots & A_{nn} \times B_{nn} \end{pmatrix} \quad \text{and} \quad A \dot{\diagup} B = \begin{pmatrix} \frac{A_{11}}{B_{11}} & \cdots & \frac{A_{1n}}{B_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{A_{n1}}{B_{n1}} & \cdots & \frac{A_{nn}}{B_{nn}} \end{pmatrix}$$

then the covariance matrix of X across years is computed as

$$C = \sum_{i=1}^{n} x_i x_i' \dot{\diagup} D$$

where D is defined as

$$D = \sum_{i=1}^{n} d_i d_i'$$

The ratio used to estimate covariances can also be written as

$$\begin{pmatrix} \frac{C_{1,1}}{D_{1,1}} & \frac{C_{1,2}}{D_{1,2}} & \cdots & \frac{C_{1,T}}{D_{1,T}} \\ \frac{C_{2,1}}{D_{2,1}} & \frac{C_{2,2}}{D_{2,2}} & & \vdots \\ \vdots & & \ddots & \\ \frac{C_{T,1}}{D_{T,1}} & & & \frac{C_{T,T}}{D_{T,T}} \end{pmatrix}$$

where each entry in the matrix C is divided by the corresponding entry in D.

In order to perform tests on the estimated covariances we first define a vector m composed of the elements of the covariance matrix C, that is $m = vech\,(C)$.[61]

Let m_i be the distinct elements of the individual cross product matrix $x_i x_i'$ and let p_i be a vector

$$p_i = \begin{pmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{iT} \end{pmatrix}$$

whose elements take value 1 if the corresponding element in the vector m_i is different from zero and value 0 otherwise. We can then define a $(T \times T)$ matrix

$$P = \sum_{i=1}^{n} p_i p_i'$$

which tells us how many non zero elements are summed within each cell of the Q matrix of uncorrected fourth moments of vector x_i

$$Q = \sum_{i=1}^{n} m_i m_i{}'$$

If we let the vector of deviations $(m_i - m)$ have zero elements whenever m_i is zero, the cross-sectional variance of the individual cross-products m_i is estimated as

$$V = \sum_{i=1}^{n} (m_i - m)(m_i - m)' \dot{\diagup} P$$

Notice that Q and V are related by $V = Q \dot{\diagup} P - mm'$. Define now $S = Vech\,(D)$. Under general conditions (Chamberlain, 1983, 1984), independence of the x_i implies that the sample mean of m_i has an asymptotic normal distribution

$$\sqrt{S} \dot{\times} (m - \mu) \sim^a N(0, V^*)$$

where μ is the expectation of m_i (that is the true covariance value) and $V^* = E(m_i m_i') - E(m_i)\,E(m_i')$.

In finite samples we can approximate V^* by using the estimated $\hat{V}$, which is the empirical equivalent.

Finally we define

$$\check{N} = \sqrt{S}\sqrt{S}'$$

and we use $\hat{V}$ to construct a new matrix $U = \hat{V} \dot{\diagup} \check{N}$. A typical element of the matrix U is $U_{uv} = cov\,(m_u, m_v)$. Here $m_u = cov\,(x_{it} x_{it-k})$ and $m_v = cov\,(x_{is} x_{is-j})$ are both

[61] Since C is symmetric we know that m has $T\,(T+1)\,/2$ elements.

elements of the m vector. The asymptotic standard error of the element m_u of m is given by $(U)^{1/2}$.

Furthermore, it can be shown that, for

$$m'U^{-1}m \sim^a \chi^2_{(\dim[m])}$$

under the null that all the elements of m are zero[62]. Of course we could select appropriate elements of m and V to run join tests of zero restrictions.

If we want to look at the covariance matrix based only on the lag of t, regardless of the specific year, we can define

$$\bar{C}_0 = \frac{\sum_{t=1}^{T} C_{t,t} D_{t,t}}{\sum_{t=1}^{T} D_{t,t}}$$

and in general

$$\bar{C}_j = \frac{\sum_{t=1}^{T} C_{t,t-j} D_{t,t-j}}{\sum_{t=1}^{T} D_{t,t-j}}$$

D.3 Estimation of the technology parameters

The Minimization Problem

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 $\Phi(\bullet)$.

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

$$\min_{\{A,B,r,s\} \in \Re^4} \left\{ \widehat{F(H)} - \left\{ A H_1^r + (1-A)\left[B H_2^s + (1-B) H_3^s\right]^{\frac{1}{s}} \right\}^{\frac{1}{r}} \right\}^2$$

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

[62] In fact, following Newey (1985) we would have that

$$mR^- m' \sim \chi^2_{(\dim[m])}$$

where R^- is a generalised inverse of $R = WUW'$ and $W = I - G(G'AG)^{-1} G'A$. The matrix A is a conformable positive definite weighting matrix (often an identity matrix) and G is normally the matrix of first derivatives of some function that is fitted to the covariance structure and has dimension $\dim[m] \times p^*$ where p^* is the number of parameters in the fitted function. In our case however p^* is equal to zero, which makes G a matrix with dimension $\dim[m] \times 0$ (or we can think of it as a matrix consisting of zeroes only). Therefore $W = I - G(G'AG)^{-1} G'A = I$ and we get the distribution result.

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

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

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[63]; (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).

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

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 conditions 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 $E\left(\{H_1, H_2, H_3\}' Z\right) \neq 0$.

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.[64] 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, 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

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

[64] 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.

where $\wp_Z = Z(Z'Z)^{-1} Z'$ 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 $\Sigma = E(RR')$ 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 $E(R'Z) = 0$ 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 $COV\left(\widehat{e}_i, \widehat{e}_j\right) = 0$ when $i \neq j$ and we have that

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

which can be written in more extensive form as

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

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

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

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

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 ?