Sources of Lifetime Inequality - AER - 2011

Wenzhi Wang [*]

June 9, 2023

Introduction

We view lifetime inequality through the lens of a risky human capital model. Agents differ in terms of three initial conditions: initial human capital, learning ability, and financial wealth. Initial human capital can be viewed as controlling the intercept of an agent's mean earnings profile, whereas learning ability acts to rotate this profile. Human capital and labor earnings are risky, as human capital is subject to idiosyncratic shocks each period.

We exploit the fact that the model implies that late in life little or no new human capital is produced. As a result, moments of the change in wage rates for these agents are almost entirely determined by shocks, rather than by shocks and the endogenous response of investment in human capital to shocks and initial conditions. We therefore estimate the shock process using precisely these moments for older males in US data.

We stress an important caveat in interpreting results on the importance of variations in initial conditions. The distribution of initial conditions at a specific age is an endogenously determined distribution from the perspective of an earlier age. This article is silent on the prior forces which shape the individual differences that we analyze at age 23.

1. The Model

An agent maximized expected lifetime utility, taking initial financial wealth k_1, initial human capital h_1, and learning ability a as given. The decision problem for an agent born at time t is stated below.

$$\max_{\{c_j,l_j,s_j,h_j,k_j\}_{j-1}} \mathbb{E}\left[\sum_{j=1}^{J}\beta^{j-1}u\left(c_j\right)\right] \tag{1}$$

subject to

$$c_j + k_{j+1} = k_j\left(1+r_{t+j-1}\right) + e_j - T_{j,t+j-1}\left(e_j, k_j\right), \forall j, k_{J+1} = 0 \tag{2a}$$

[*] This note is written down during my M.phil. period at the University of Oxford.

$$e_j = R_{t+j-1} h_j l_j \text{ if } j < J_R \text{ , and } e_j = 0 \text{ , otherwise} \tag{2b}$$

$$h_{j+1} = \exp\left(z_{j+1}\right) H\left(h_j, s_j, a\right) \text{ and } l_j + s_j = 1, \forall j \tag{2c}$$

The only source of risk to an agent over the working lifetime comes form idiosyncratic shocks to an agent's human capital. Let $\boldsymbol{z}^j = (z_1, \ldots, z_j)$ denote the j-period history of these shocks. Thus, the optimal consumption choice $c_{j,t+j-1}(\boldsymbol{x}_1, \boldsymbol{z}^j)$ for an age j agent at time $t + j - 1$ is risky as it depends on shocks $\boldsymbol{z}^j$ as well as initial conditions $\boldsymbol{x}_1 = (h_1, k_1, a)$. The period budget constraint says that consumption c_j plus financial asset holding k_{j+1} equals earnings e_j plus the value of assets $k_j(1 + r_{t+j-1})$ less net taxes $T_{j,t+j-1}$. Financial assets pay a risk-free, real return r_{t+j-1} at time $t + j - 1$. Earnings e_j before a retirement age J_R equal the product of a rental rate R_{t+j-1} for human capital services, an agent's human capital h_j, and the fraction l_j of available time put into market work. Earnings are zero at and after the retirement age J_R. An agent's future human capital h_{j+1} is an increasing function of an idiosyncratic shock z_{j+1}, current human capital h_j, time devoted to human capital or skill production s_j, and an agent's learning ability a. Learning ability is fixed over an agent's lifetime and is exogenous.

We now embed this decision problem within a general equilibrium framework and focus on balanced-growth equilibria. There is an aggregate production function $F(K_t, L_t A_t)$ with constant returns in aggregate capital and labor (K_t, L_t) and with labor augmenting technical change $A_{t+1} = A_t(1 + g)$. Aggregate variables are sums of the relevant individual decisions across agents. In defining aggregates, ϕ is a time-invariant distribution over initial conditions $\boldsymbol{x}_1$, and μ_j is the fraction of age j agents in the population. Population fractions satisfy $\sum_{j=1}^{J} \mu_j = 1$ and $\mu_{j+1} = \mu_j/(1 + n)$, where n is a constant population growth rate. In the analysis of equilibrium, we consider the case where initial financial assets are zero and, thus, ϕ is effectively a bivariate distribution over $\boldsymbol{x}_1 = (h_1, a)$.

$$\begin{aligned}
K_t &\equiv \sum_{j=1}^{J} \mu_j \int \mathbb{E}\left[k_{j,t}\left(\boldsymbol{x}_1, \boldsymbol{z}^j\right)\right] d\psi \text{ and } L_t \equiv \sum_{j=1}^{J} \mu_j \int \mathbb{E}\left[h_{j,t}\left(\boldsymbol{x}_1, \boldsymbol{z}^j\right) l_{j,t}\left(\boldsymbol{x}_1, \boldsymbol{z}^j\right)\right] d\psi \\
C_t &\equiv \sum_{j=1}^{J} \mu_j \int \mathbb{E}\left[c_{j,t}\left(\boldsymbol{x}_1, \boldsymbol{z}^j\right)\right] d\psi \text{ and } T_t \equiv \sum_{j=1}^{J} \mu_j \int \mathbb{E}\left[T_{j,t}\left(e_{j,t}, k_{j,t}\right)\right] d\psi
\end{aligned} \tag{3}$$

Definition 1. A balanced-growth equilibrium is a collection of decisions $\left\{\left\{c_{j,t}, l_{j,t}, s_{j,t}, h_{j,t}, k_{j,t}\right\}_{j=1}^{J}\right\}_{t=-\infty}^{\infty}$, factor prices, government spending and taxes $\{R_t, r_t, G_t, T_t\}_{t=-\infty}^{\infty}$, and a distribution ϕ over initial conditions such that

1. Agent decisions are optimal, given factor prices.
2. Competitive Factor Prices: $R_t = A_t F_2\left(K_t, L_t A_t\right)$ and $r_t = F_1\left(K_t, L_t A_t\right) - \delta$.
3. Resource Feasibility: $C_t + K_{t+1}(1 + n) + G_t = F\left(K_t, L_t A_t\right) + K_t(1 - \delta)$.
4. Government Budget: $G_t = T_t$.
5. Balanced Growth: (i) $\left\{c_{j,t}, k_{j,t}\right\}_{j=1}^{J}$ grow at rate g as a function of time, whereas $\left\{l_{j,t}, s_{j,t}, h_{j,t}\right\}_{j=1}^{J}$ are time invariant. (ii) (G_t, T_t, R_t) grow at rate g, whereas r_t is time invariant.

Our focus on balanced-growth equilibria requires that individual decisions, aggregate variables, and factor prices grow at constant rates. Balanced growth leads us to employ homothetic preferences and a constant returns technology. More specifically, we use the property that if preferences over lifetime consumption plans are homothetic and the budget set for consumption plans is homogeneous of degree 1 in rental rates, then optimal consumption plans are homogeneous of degree 1 in rental rates. [1]

The functional forms that we employ are provided below. The equilibrium concept does not restrict the functional forms for the human capital production function $H(h,s,a)$, the distribution of initial condition ϕ, or the nature of idiosyncratic shocks. The human capital production function is of the Ben-Porath class which is widely used in empirical work. The distribution ϕ is a bivariate lognormal distribution which allows for a skewed distribution of initial human capital. Recall that our equilibrium analysis considers the case where initial assets are set to zero. Idiosyncratic shocks are independent and identically distributed over time and follow a normal distribution:

$$u(c) = \frac{c^{1-\rho}}{1-\rho} \tag{4a}$$

$$F(K,LA) = K^{\gamma}(LA)^{1-\gamma} \tag{4b}$$

$$H(h,s,a) = h + a(hs)^{\alpha} \tag{4c}$$

$$\boldsymbol{x} = (h_1, a) \sim \phi = \text{LogNormal}(\boldsymbol{\mu}_x, \boldsymbol{\Gamma}) \tag{4d}$$

$$z \sim \text{Normal}(\mu, \sigma^2). \tag{4e}$$

We comment on four key features of the model. First, while the earnings of an agent are stochastic, the earnings distribution for a cohort of agents evolves deterministically. This occurs because the model has idiosyncratic but no aggregate risk. Second, the model has two sources of growth in earnings dispersion within a cohort–agents have different learning abilities and different shock realizations. The next section characterizes empirically the rise in US earnings dispersion over the working lifetime. Third, although the model has a single source of shocks, which are independently and identically distributed over time, we will show that this structure is sufficient to endogenously produce many of the statistical properties of earnings that researchers have previously estimated.Fourth, the model implies that the nature of human capital shocks can be identified from wage rate data, independently from all other model parameters. This holds, as an approximation, because the model implies that the production of human capital goes to zero towards the end of the working lifetime.

[1] Let $\Lambda(\boldsymbol{x}_1, \widetilde{\boldsymbol{R}})$ denote the set of lifetime consumption plans satisfying budget conditions (2a)-(2c), given initial conditions $\boldsymbol{x}_1$ and rental rates $\widetilde{\boldsymbol{R}} = (R_1, \ldots, R_J)$. $\Lambda(\boldsymbol{x}_1, \widetilde{\boldsymbol{R}})$ is homogeneous in $\widetilde{\boldsymbol{R}}$ provided $\boldsymbol{c} \in \Lambda(\boldsymbol{x}_1, \widetilde{\boldsymbol{R}}) \implies \lambda\boldsymbol{c} \in \Lambda\left(\boldsymbol{x}_1, \lambda\widetilde{\boldsymbol{R}}\right), \forall \lambda > 0$. $\Lambda(\boldsymbol{x}_1, \widetilde{\boldsymbol{R}})$ has this property when taxes T_{jt} are homogeneous of degree 1 in earnings and assets and when initial assets are zero. The model tax system induces this property when $T_{jt}(\widetilde{\boldsymbol{R}}, h_j, l_j, k_j)$ is homogeneous of degree 1 in $(\widetilde{\boldsymbol{R}}, k_j)$.

2. Empirical Analysis

We use data to address two issues. First, we characterize how mean earnings and measures of earnings dispersion and skewness evolve with age for a cohort. Second, we estimate a human capital stock process from wage rate data.

2.1. Age Profiles

We estimate age profiles for mean earnings and measures of earnings dispersion and skewness for age 23 to 60. We use earnings data for males who are the head of the household from the PSID 1969-2004 family files. To calculate earnings statistics at a specific age and year, we employ a five-year age bin. For males over age 30, we require that they work between 520 and 5820 hours per year and earn at least \$1500 (in 1986 prices) a year. For males age 30 and below, we require that they work between 260 and 5820 hours per year and earn at least \$1000 (in 1986 prices).

These selection criteria are motivated by several considerations. First, the PSID has many observations in the middle but relatively fewer at the beginning or end of the working life cycle. By focusing on ages 23-60, we have at least 100 observations in each age-year bin with which to calculate earnings statistics. Second, labor force participation falls near the traditional retirement age for reasons that are abstracted from in the model. This motivates the use of a terminal age that is below the traditional retirement age. Third, the hours and earnings restrictions are motivated by the fact that within the model the only alternative to time spent working is time spent learning. For males above 30, the minimum hours restriction amounts to a quarter of full-time work hours, and the minimum earnings restriction is below the annual earnings level of a full-time worker working at the federal minimum wage. For younger males, we lower both the minimum hours and earnings restrictions to capture students doing summer work or working part time while in school.

We now document how mean earnings, two measures of earnings dispersion, and a measure of earnings skewness evolve with age for cohorts. We consider two measures of dispersion: the variance of log earnings and the Gini coefficient of earning. We measure skewness by the ratio of mean earnings to median earnings.

The methodology for extracting age effects is in two parts. First, we calculate the statistic of interest for males in age bin j at time t and label this $stat_{j,t}$. For example, for mean earnings we set $stat_{j,t} = \ln(e_{jt})$, where e_{jt} is real mean earnings of all males in the age bin centered at age j in year t. Second, we posit a statistical model governing the evolution of the earnings statistic as indicated below. The earnings statistic is viewed as being generated by several factors that we label cohort (c), age (j), and time (t) effects. We wish to estimate the age effects β_j^{stat}. We employ a statistical model, as our economic model is not sufficiently rich to capture all aspects of time variation in the data:

$$\text{stat}_{j,t} = \alpha_c^{\text{stat}} + \beta_j^{\text{stat}} + \gamma_t^{\text{stat}} + \epsilon_{j,t}^{\text{stat}}.$$

The linear relationship between time t, age j, and birth cohort $c = t - j$ limits the applicability of this regression specification. Specifically, without further restrictions the regressors in this system are colinear, and these effects cannot be estimated. This identification problem is well known. Any trend in the data can be arbitrarily reinterpreted as due to year (time) effects or alternatively as due to age or cohort effects.Given this problem, we provide two alternative measures of age effects. These correspond to the cohort effects view where we set $\gamma_t^{\text{stat}} = 0, \forall t$ and the time effects view where we set $\alpha_c^{\text{stat}}, \forall c$. We use OLS to estimate the coefficients.

We will ask the economic model to match both views of the evolution of the earnings distribution. Given the lack of a consensus in the literature, we are agnostic about which view should be stressed. To conserve space, the article highlights the results of matching the time effects view in the main text but summarizes results for the cohort effects view later.

2.2. Human Capital Shocks

The model implies that an agent's wage rate, defined within the model as earnings per unit of work time, equals the product of the rental rate and an agent's human capital. Here it is important to recall that within the model work time and learning time are distinct activities. The model also implies that late in the working lifetime human capital investment are approximately zero. This occurs as the number of working periods over which an agent can reap the returns to these investments falls as the agent approaches retirement. The upshot is that when there is no human capital investment over a period of time, then the change in an agent's wage rate is in theory entirely determined by rental rates and the human capital shock process and not by any other model parameters.

In what follows, assume that in periods t through $t+n$ an individual devotes zero time learning. The first equation below states that the wage w_{t+n} is determined by the rental rate R_{t+n}, shocks $(z_{t+1}, \ldots, z_{t+n})$, and human capital h_t. Here it is understood that $h_{t+1} = \exp(z_{t+1}) H(h_t, s_t, a) = \exp(z_{t+1}) \left[h_t + f(h_t, s_t, a) \right]$ and that there is zero human capital production in periods when there is no investment (i.e., $f(h, s, a) = 0$ when $s = 0$). The second equation takes logs of the first equation, where a hat denotes the log of a variable. The third equation states that measured n-period log wage differences (denoted $y_{t,n}$) are true log wage differences plus measurement error differences $\varepsilon_{t+n} - \varepsilon_t$. The third equation highlights the point that log wage differences are due solely to rental rate differences and shocks:

$$w_{t+n} = R_{t+n} h_{t+n} = R_{t+n} \exp(z_{t+n}) H(h_{t+n-1}, 0, a) = R_{t+n} \prod_{i=1}^{n} \exp(z_{t+i}) h_t$$

$$\widehat{w}_{t+n} = \ln w_{t+n} = \widehat{R}_{t+n} + \sum_{i=1}^{n} z_{t+i} + \widehat{h}_t$$

$$y_{t,n} = \widehat{w}_{t+n} - \widehat{w}_t + \epsilon_{t+n} - \epsilon_t = \widehat{R}_{t+n} - \widehat{R}_t + \sum_{i=1}^{n} z_{t+i} + \epsilon_{t+n} - \epsilon_t.$$

Our strategy for estimating the nature of human capital shocks is based on the log-wage-difference equation. Thus, it is important to be able to measure the wage concept used in the theory and to have individuals for which the assumption of no time spent accumulating human capital is a reasonable approximation.. The wage concept in the theoretical model is earnings per unit of work time. Thus, two critical assumptions are that (i) measured work time is only work time and not a combination of work and learning time - distinct activities in human capital theory and in our model - and (ii) no time on the job or off the job is spent learning and, thus, producing new human capital. We focus on older workers to address both of these issues. Young workers are likely, in our view, to be problematic on both issues.

We use the log-wage-difference equation and make some specific assumptions. We assume that both human capital shocks z_t and measurement errors ε_t are independent and identically distributed over time and people. Furthermore, we assume that $z_t \sim N(\mu, \sigma^2)$ and $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$. These assumptions imply the three cross-sectional moment conditions below:

$$
\begin{aligned}
E\left[y_{t,n}\right] &= \widehat{R}_{t+n} - \widehat{R}_t + n\mu \\
\text{var}\left(y_{t,n}\right) &= n\sigma^2 + 2\sigma_\epsilon^2 \\
\text{cov}\left(y_{t,n}, y_{t,m}\right) &= m\sigma^2 + \sigma_\epsilon^2 \text{ for } m < n
\end{aligned}
$$

We calculate real wages in PSID data as total male labor earnings divided by total hours for male head of household, using the Consumer Price Index to convert nominal wages to real wages. We follow males for four years and thus calculate three log wage differences (i.e., $y_{t,n}$ for $n = 1,2,3$). In utilizing the wage data we impose the same selection restriction as in the construction of the age-earnings profiles but also exclude observations for which earnings growth is above (below) 20 (1/20) to trim potential extreme measurement errors. In estimation we use all cross-sectional variances and all cross-sectional covariances aggregated across panel years. For each year, we generate the sample analog to the moments,

$$
\mu_{t,n} \equiv \frac{1}{N_t}\sum_{i=1}^{N_t} y_{t,n}^i \text{ and } \frac{1}{N_t}\sum_{i=1}^{N_t}\left(y_{t,n}^i - \mu_{t,n}\right)^2 \text{ and } \frac{1}{N_t}\sum_{i=1}^{N_t}\left(y_{t,n}^i - \mu_{t,n}\right)\left(y_{t,m}^i - \mu_{t,m}\right).
$$

We stack the moments across the panel years and use a two-step GMM estimation with an identity matrix as the initial weighting matrix.

3. Setting Model Parameters

The parameters are listed in Table 2 and are set in two steps. The first collection of model parameters is set without solving the model. The remaining model parameters are set so that the equilibrium properties of the model best match the earnings distribution facts documented in the pervious section while matching some steady-state quantities.

The remaining model parameters are set so that the equilibrium properties of the model best match the earnings distribution facts. The parameters selected are those governing the distribution of initial conditions $\psi = LN(\boldsymbol{\mu_x}, \boldsymbol{\Gamma})$, the elasticity of the human capital production function α, and the agent's discount factor β.

Category	Symbol	Parameter value
Demographics	(J, J_R, n)	$(J, J_R, n) = (53, 39, 0.012)$
Preferences	$\beta, u(c) = c^{(1-\rho)}/(1-\rho)$	$(\beta, \rho) = (0.981, 2)$
Technology	(γ, δ, g)	$(\gamma, \delta, g) = (0.322, 0.067, 0.0019)$
Tax system	$T_j = T_j^{ss} + T_j^{inc}$	$T_j^{ss}(e_j) = 0.106 e_j$ for $j < J_R$ $T_j^{ss}(e_j) = -0.4\bar{e}$ otherwise T_j^{inc}—see text
Human capital shocks	$z \sim N(\mu, \sigma^2)$	$(\mu, \sigma) = (-0.029, 0.111)$
Human capital technology	$h' = \exp(z')H(h, s, a)$ $H(h, s, a) = h + a(hs)^{\alpha}$	$\alpha = 0.70$
Initial conditions	$\psi = LN(\boldsymbol{\mu}_{\mathbf{x}}, \Sigma)$	$\boldsymbol{\mu}_{\mathbf{x}} = (\mu_h, \mu_a) = (4.66, -1.12)$ $(\sigma_h^2, \sigma_a^2, \sigma_{ha}) = (0.213, 0.012, 0.041)$

Figure 1: Table 2: Parameter Values

4. Properties of the Benchmark Model

4.1. Dynamics of the Earnings Distribution

The model generates the hump-shaped earnings profile for a cohort by a standard human capital argument. Early in the working life cycle, individuals devote more time to human capital production than at later ages. These time allocation decisions lead to a net accumulation of human capital in the early part of the working life cycle. Thus, mean earnings increase with age as human capital and mean time worked increase with age.

Toward the end of the working life cycle, mean human capital levels fall. This happens as the mean multiplicative shock to human capital is smaller than one (i.e., $\mathbb{E}\left[\exp(z)\right] = \exp\left(\mu + \sigma^2/r\right) < 1$). This corresponds to the notion that on average human capital depreciates. The implication is that average earnings fall later in life because growth in the rental rate of human capital is not enough to offset the mean fall in human capital.

The mean human capitla profile is hump-shaped and that it is flatter than the earnings profile. A relatively flat mean human capital profile and a declining time allocation profile to human capital production is how the model accounts for a hump-shaped earnings profile. The fact that the mean human capital profile is flatter than the earnings profile means that average human capital as of age 23 is quite high. This is a key reason why we find in the next section that human capital differences are such an important source of individual differences at age 23 compared to ability differences.

Two forces account for the rise in earnings dispersion. First, since individual human capital is repeatedly hit by shocks, these shocks are a source of increasing dispersion in human capital and earnings as a cohort ages. Second, differences in learning ability across agents produce mean earnings profiles with different slopes. This follows since within an age group, agents with high learning ability choose to produce more human capital and devote more time to human capital production than their low ability counterparts.

4.2. Earnings Dispersion: Risk versus Ability Differences

We now try to understand the quantitative importance of risk and ability differences for producing the increase in earnings dispersion in the benchmark model. We do so by either eliminating ability differences or eliminating shocks. The analysis holds factor prices constant as risk or ability differences are varied.

We eliminate ability differences by changing the initial distribution so that all agents have the same learning ability, which we set equal to the median ability. In the process of changing learning ability, we do not alter any agent's initial human capital.

To highlight the role of human capital risk, we eliminate idiosyncratic risk by setting $\sigma = 0$. We adjust the mean log shock μ to keep the mean shock level constant. We do not change the distribution of initial conditions. Eliminating risk results in substantial changes in the time allocation decisions of agents with relatively high learning ability. Absent risk, these agents allocate an even larger fraction of time into human capital accumulation. This leads to very high earnings dispersion early in life as some of these agents have very low earnings. Later in life these agents have higher earnings than agents with lower learning ability, other things equal.

4.3. Properties of the Initial Distribution

A key finding is that human capital and learning ability are positively correlated at age 23.

It also summarizes the distribution of initial conditions in the model with initial wealth differences. We choose this distribution to be in essence a trivariate lognormal distribution. The parameters related to financial wealth are set to match features of financial wealth holding of young households in US data as is explained in the next section. The remaining parameters of this distribution are selected to match the earnings facts in Figure 1.

4.4. Statistical Models of Earnings