

# Permanent Income Inequality

PRELIMINARY AND INCOMPLETE

**Brant Abbott**

Queen's University

abbottbrant@gmail.com

**Giovanni Gallipoli**

University of British Columbia

gallipol@mail.ubc.ca

THIS DRAFT: FEBRUARY 27, 2018

**Abstract.** We estimate the level and evolution of inequality in assets, human capital wealth and permanent income. Our definition of the latter variable does not rely on a specific utility function and imposes no restrictions on income processes. We characterize the distribution of human wealth using nonparametric identification results that allow for state-dependent stochastic discounting and unobserved heterogeneity. Accounting for the value of human capital delivers a different view of inequality. We find that (i) in 2016 the top 10% shares of total wealth and permanent income were roughly 1/3 lower than the corresponding share of assets; (ii) between 1989 and 2016 the top 10% shares of total wealth and permanent income grew significantly faster than the corresponding share of asset wealth. Hence, human wealth has had a mitigating influence on overall inequality but this mitigating effect has declined over time. We show that households at the top of the assets distribution have not increased their share of human wealth. Instead, higher concentration of permanent income is due to the growing importance of assets in lifetime wealth portfolios.

**JEL Classification:** D3, G1, I24, E21, J01

**Keywords:** Inequality, Earnings, Wealth, Human Capital

# 1 Introduction

*“The magnitudes termed ‘permanent income’ and ‘permanent consumption’ that play such a critical role in the theoretical analysis cannot be observed directly for any individual consumer unit. The most that can be observed are actual receipts and expenditures during some finite period....The theoretical constructs are ex ante magnitudes; the empirical data are ex post. Yet in order to use the theoretical analysis to interpret empirical data, a correspondence must be established between the theoretical constructs and the observed magnitudes.” M. Friedman (1957)*

A primary objective of inequality research is to understand the forces shaping differences in the economic wellbeing of individuals and households. Empirical research has made progress towards this goal by analyzing inequality of observable variables, primarily income and wealth.<sup>1</sup> However, a broader assessment of economic inequality would require that one also accounts for the heterogeneity associated with future earnings potential. This is apparent in the optimal redistribution branch of the literature where equalization of marginal utilities from consumption is often assumed to be the underlying policy goal, and optimal policies depend on the unobservable value of ex-ante expected future earnings.<sup>2</sup>

Yet-to-be realized earnings may constitute the most important determinant of economic wellbeing for many households. A young person with a steeply increasing expected earnings profile may be better off than inferred by simply measuring their current net worth or income. The extent to which future earnings matter depends on how much they are discounted. Appropriate discounting of future earnings effectively accounts for the ease with which con-

---

<sup>1</sup>See for example the work of Saez and Kopczuk (2004), Piketty and Saez (2006), Saez and Zucman (2014), Bricker, Henriques, Krimmel *et al.* (2016), Kaymak and Poschke (2016) and Rios-Rull, Kuhn *et al.* (2016). An extensive literature on the distribution of wages and earnings documents widening inequality in the working population (see for example Levy and Murnane (1992), Gottschalk, Moffitt, Katz *et al.* (1994), Goldin and Katz (2007) and Autor, Katz, and Kearney (2008)). Studies of the wealth distribution focus on the financial/real wealth held by the wider population, including the unemployed and those who do not participate in the labor market (see Saez and Zucman, 2014; Bricker, Henriques, Krimmel *et al.*, 2016). More recently, the work of De Nardi, Fella, and Paz-Pardo (2016) illustrates how rich income processes (as those described in Guvenen, Karahan, Ozkan *et al.*, 2016) may be reflected in the equilibrium distribution of wealth. Interestingly, Athreya, Ionescu, and Neelakantan (2015; 2017) provide evidence that human capital investments may have important effects on financial portfolios in the cross-section and throughout the life-cycle.

<sup>2</sup>This is an extensive literature. The New Dynamic Public Finance part of the literature is surveyed by Golosov, Tsyvinski, Werning *et al.* (2006) and Kocherlakota (2010). Examples from the Ramsey planning section of the literature include Conesa, Kitao, and Krueger (2009) and Davila, Hong, Krusell *et al.* (2012).

sumption can be shifted across time periods as well as for uncertainty about future earnings and consumption.<sup>3</sup> Being constrained by a credit limit or facing a great deal of risk reduces a household's valuation of their future earnings.

In this paper we estimate pecuniary measures that reflect the values of both human capital and asset wealth. Then, we study trends in the concentration of these variables over the period 1989-2016. At the heart of our analysis are nonparametric estimates of the value to individuals of their yet-to-be realized earnings, which we refer to as their human wealth. These estimates differ from the simple expected present value of future earnings in several ways. Importantly, they feature state-dependent stochastic discounting, rather than risk-free discounting. Combining human wealth estimates with observed asset wealth data allows us to estimate lifetime wealth, which is the sum of human wealth and asset wealth. We also construct estimates of permanent income, which is the (age-adjusted) annuity value of lifetime wealth. The latter statistic is reminiscent of 'permanent income' as defined by Friedman (1957), with the obvious difference that in Friedman's model human wealth is the risk-free present value of expected future earnings. We find that in 2016 the top 10% share of permanent income was almost 1/3 lower than the top 10% share of asset wealth, and the top 10% share of lifetime wealth was even lower. However, between 1989 and 2016, permanent income concentration has grown at a yearly rate of almost 1%, much faster than the 0.5% growth in asset wealth concentration. Concentration in total (lifetime) wealth inequality has grown even faster, at almost 1.3% per year. Hence, we infer that (i) human capital has had a mitigating influence on the level of overall inequality; and (ii) this mitigating influence appears to be declining over time.

To obtain our estimates we combine data from the Panel Study of Income Dynamics (PSID) and the Survey of Consumer Finances (SCF). The PSID is useful for its panel data

---

<sup>3</sup>The way future income is discounted is important. Huggett and Kaplan (2016) convincingly argue that the true value of human capital is far below the value that would be implied by discounting future net earnings at the risk-free interest rate, an approach that is commonly advocated because of its simplicity (see Becker, 1975; Jorgenson and Fraumeni, 1989; R. Haveman and Schwabish, 2003). Mechanically discounting income flows to approximate human capital rules out state-dependent valuations of future earnings.

on earnings and consumption, which are required for identification of nonparametric human wealth valuation functions. We then apply these estimated functions to SCF data, where the resulting estimates of human wealth can be combined with observed net worth. This allows one to obtain more accurate estimates of lifetime wealth and permanent income. We do not make assumptions about the processes that generate risk in the labor market. Any aggregate risk present in the data is accounted for in our estimates of human wealth. To obtain a long enough sample to identify the aggregate risk component, we impute consumption in the PSID prior to 1999 using the method suggested by Attanasio and Pistaferri (2014). Thus, all PSID data from 1968 to 2016 are utilized when estimating human wealth.

Crucially, our estimates of human wealth account for state-dependent stochastic discount factors and for changes in marginal utility. Rather than assuming specific functional forms we estimate stochastic discount factors nonparametrically. This dispenses with several restrictions and lets data guide the choice of utility function in a flexible way. Nonparametric identification of the marginal utility function is achieved by using and extending key results in Escanciano, Hoderlein, Lewbel *et al.* (2016). This involves writing the intertemporal Euler equation in such a manner that the estimated marginal utility function is the solution of a homogeneous Fredholm equation of the second kind. Given identification of the stochastic discount factor, human wealth then depends on an integral over its possible future values multiplied by the realizations of the stochastic discount factor.

Having obtained a marginal utility function, the estimated human wealth valuation equation turns out to be the solution of an *inhomogeneous* Fredholm equation of the second kind. The non-homogeneous form of this specification demands that we extend existing results to prove nonparametric identification of human wealth.

A separate issue arises from the fact that only one realization of the future state of the world is observed for each person and time-period in our sample. Hence we do not observe the entire distribution of possible future outcomes, on which an individual's human wealth depends. We

address this data limitation by approximating the distribution of possible outcomes using those observed for individuals who are, in a way made clear later, ex-ante similar. This approach works under an identification assumption, which we refer to as conditional equivalence of expectations. This assumption simply states that individuals who are ex-ante equivalent, in terms of individual characteristics and the aggregate state, face the same distribution of ex-post outcomes. Our implementation allows for the distribution of ex-post outcomes to vary with both observable characteristics and unobservable types. Unobservable heterogeneity is potentially very important in this situation because certain forms of heterogeneity, such as heterogeneous income profiles, could lead to differences in the distributions of ex-post outcomes even if individuals have identical ex-ante observable characteristics. To identify unobservable types we adapt the method developed by Bonhomme, Lamadon, and Manresa (2017) in such a way that the number of unobservable types is chosen to reflect the degree of ex-ante heterogeneity in the sample. Inclusion of these types in the conditioning set assuages our concern that unobserved differences in human wealth may lead to underestimates of the degree of inequality.

The upshot of our econometric work is an analysis of inequality of human wealth, lifetime wealth and permanent income that can be immediately related to many existing studies of inequality based on observed incomes or net worth. This approach allows one to ask questions like ‘what is the top 1% share of lifetime wealth?’ and ‘how has the Gini coefficient of permanent income changed over time?’, despite the fact that both lifetime wealth and permanent income are ex-ante magnitudes that cannot be directly observed. These variables are closely related to economic wellbeing, and certainly more so than current income or assets alone. By their nature, these theoretical constructs are identified through a set of structural assumptions, hence the usefulness of our estimates is limited by the plausibility of those assumptions. Our use of nonparametric methods ensures that only the low level assumptions of the theory, such as utility maximization, are used to identify the value of human wealth, rather than higher level assumptions, such as specific utility functional forms or wage generating processes. As

such, we make the assumptions underlying our estimates as plausible as possible, while still maintaining comparability between our analysis and existing studies of inequality based on observable wealth and income.

## 2 Theory

### 2.1 Lifetime Wealth

The state of the economy at time  $t$  is represented by  $\Omega_t$ . The history of states of the world is then  $\Omega^t = \{\Omega_0, \Omega_1, \dots, \Omega_t\}$ .  $\Omega_t$  includes realizations of all aggregate and idiosyncratic (individual-level) risk. An individual's observable characteristics, such as education, age and gender, are contained in the vector  $X_{it}$ . An individual's unobservable type, which may be informative about their expected earnings or consumption profile, is denoted by  $\eta_i$ . If an individual is married they will have a spouse with observable characteristics  $X_{jt}$  and unobservable type  $\eta_j$ . A household's wealth portfolio is a vector containing various assets and liabilities. For an unmarried household this vector is  $\mathbf{a}_{it} = \{a_{it}^\kappa\}_{\kappa \in k}$ , where  $a_{it}^\kappa$  is the individual's position in asset  $\kappa$ . For a married household consisting of an individual  $i$  and their spouse  $j$ , the wealth portfolio is  $\mathbf{a}_{(ij)t}$ .

**Individual Value Functions.** An individual enjoys utility from consumption and leisure, denoted  $u(c_{it}, \ell_{it})$ , and (possibly) from being married to their spouse, denoted  $\heartsuit_{it(j)}$ . An individual's value function when single,  $V_i^S$ , depends on their own state variables and their beliefs about marital prospects. The value function when married,  $V_i^M$ , depends on both own and spousal state variables, and beliefs about the prospect of remaining married. An individual may supply a fraction  $h_{it}$  of their time in the labour market, for which they earn a wage  $w_{it}$ . Wages vary with  $X_{it}$  and  $\Omega_t$ .

If individual  $i$  is single at time  $t$  their value function  $V_i^S$  will depend on a continuation

value at time  $t + 1$  that includes the possibilities of choosing to get married or remain single in the following period:

$$\begin{aligned}
V_i^S(\mathbf{a}_{it}, X_{it}, \eta_i, \Omega^t) = & \max_{c_{it}, \ell_{it}, h_{it}, \mathbf{a}_{it+1}} \left\{ u(c_{it}, \ell_{it}) \right. \\
& + \beta(1 - \mu_{it}) E_{\{\Omega^{t+1}\}} [V_i^S(\mathbf{a}_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1})] \\
& \left. + \beta \mu_{it} E_{\{\Omega^{t+1}, X_{jt+1}, \eta_j, \mathbf{a}_{jt+1}\}} [V_i^M(\mathbf{a}_{(ij)t+1}, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1})] \right\}. \tag{1}
\end{aligned}$$

The probability  $\mu_{it} = \mu(X_{it}, \eta_i, \Omega^t)$  is the conditional probability that  $i$  chooses to get married next period, after meeting potential partners. This probability depends on individual characteristics and the state of the world. In the event that  $i$  chooses to marry, their indirect utility will depend on the wealth and characteristics of their partner,  $\mathbf{a}_{jt+1}$  and  $X_{jt+1}$ , as well as the state of the world next period. Thus, the expected value of being married is taken over the distribution of these variables among the  $j$  individuals that person  $i$  might choose to marry. The assets of a newly formed married household will be the sum of the spouses initial individual assets:  $\mathbf{a}_{(ij)t+1} = \mathbf{a}_{it+1} + \mathbf{a}_{jt+1}$ .

The consumption choice of  $i$  is defined over their current budget set

$$\sum_{\kappa \in k} a_{it+1}^{\kappa} + c_{it} \leq w_{it} h_{it} + \sum_{\kappa \in k} R_t^{\kappa} a_{it}^{\kappa} - T_t(\mathbf{a}_{it}, w_{it}, h_{it}), \tag{2}$$

where  $R_t^{\kappa}$  is the one-period return on asset  $\kappa$ , and  $T(\mathbf{a}_{it}, w_{it}, h_{it})$  is a function summarizing all tax liabilities. The individual's time constraint  $\ell_{it} = 1 - h_{it}$  and current borrowing constraint  $\sum_{\kappa \in k} a_{it+1}^{\kappa} \geq \underline{a}_{it}$  also affect these choices.

If individual  $i$  is married to individual  $j$  at time  $t$ , then  $i$ 's value function will include a continuation value that allows for the possibilities of staying married or separating in the

following year:

$$\begin{aligned}
V_i^M(\mathbf{a}_{(ij)t}, X_{it}, X_{jt}, \eta_i, \eta_j, \Omega^t) = & \quad (3) \\
u(c_{it}^*, \ell_{it}^*) & \\
+ \beta \{ (1 - \tilde{\mu}_{it}) E_{\{\Omega^{t+1}, \mathbf{a}_{it+1}\}} [V_i^S(\mathbf{a}_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1}) | \mathbf{a}_{(ij)t+1}^*] & \\
+ \tilde{\mu}_{it} E_{\{\Omega^{t+1}\}} [V_i^M(\mathbf{a}_{(ij)t+1}^*, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1})] \} + \heartsuit_{it(j)}. &
\end{aligned}$$

In the above equation the values  $(\mathbf{a}_{(ij)t+1}^*, c_{it}^*, \ell_{it}^*)$  are the values of household savings, as well as consumption and leisure for individual  $i$ , that result from the joint household optimization problem described below. The parameter  $\tilde{\mu}_{it} = \mu(X_{it}, X_{jt}, \eta_i, \eta_j, \Omega^t)$  is the conditional probability of a household choosing to stay married. If the household divorces before next period their asset portfolio is split and individual  $i$  receives a part  $\mathbf{a}_{it+1}$  of it. Because there may be uncertainty about the divorce settlement, a conditional expectation over possible asset divisions is taken when evaluating the divorce part of the continuation value. While we don't model the choice of getting married explicitly, we assume that the marriage shock  $\heartsuit_{it(j)}$  captures the presence of non-pecuniary returns to being married to person  $j$ . These returns are assumed to be additively separable and drop out of all marginal calculations.

**Household Planner Problem.** Once married, the joint optimization problem of the spouses can be viewed as that of a planner who maximizes a weighted average of the spouses' utilities using a set of Pareto weights. Above we have denoted by  $V_i^M$  the utility of person  $i$  when they are assigned the allocations that the household planner finds optimal. Next, we need to distinguish this from person  $i$ 's utility under (possibly) non-optimized allocations, which we

denote by  $\tilde{V}_i^M$ . The problem of the household planner is:

$$V_{(ij)}^M(\mathbf{a}_{(ij)t}, X_{it}, X_{jt}, \eta_i, \eta_j, \Omega^t) = \max_{\mathbf{b}_{(ij)t}} \left\{ \lambda_{(ij)} \tilde{V}_i^M(\mathbf{a}_{(ij)t}, X_{it}, X_{jt}, \eta_i, \eta_j, \Omega^t) + (1 - \lambda_{(ij)}) \tilde{V}_j^M(\mathbf{a}_{(ij)t}, X_{jt}, X_{it}, \eta_j, \eta_i, \Omega^t) \right\}, \quad (4)$$

where the decision vector is  $\mathbf{b}_{(ij)t} = \{c_{it}, c_{jt}, \ell_{it}, \ell_{jt}, h_{it}, h_{jt}, \mathbf{a}_{(ij)t+1}\}$ , and  $\lambda_{(ij)}$  is the Pareto weight on individual  $i$  in the household planning problem.

The feasible consumption set for married households is determined by the budget constraint

$$\sum_{\kappa \in k} a_{(ij)t+1}^{\kappa} + c_{(ij)t} \leq w_{it} h_{it} + w_{jt} h_{jt} + \sum_{\kappa \in k} R_t^{\kappa} a_{(ij)t}^{\kappa} - T_t(\mathbf{a}_{(ij)t}, w_{it}, w_{jt}, h_{it}, h_{jt}), \quad (5)$$

where  $c_{(ij)t}$  is total consumption expenditure of the household. This is related to the consumption resources allocated to each spouse by the constraint  $c_{(ij)t} = \vartheta(c_{it} + c_{jt})$ , where  $\vartheta$  represents an adult equivalence scale. Individual time allocation constraints  $\ell_{it} = 1 - h_{it}$  and  $\ell_{jt} = 1 - h_{jt}$ , and the household borrowing limit  $\sum_{\kappa \in k} a_{(ij)t+1}^{\kappa} \geq \underline{a}_{(ij)t}$  also constrain the household planner's choices.

## 2.2 Valuation of Human Wealth

We model human capital as an asset that pays dividends in the form of earnings; hence, human wealth is the value of this asset. We assume that the dividends correspond to the earnings of a worker who optimally chooses labor supply. Letting  $\theta_{it}$  be the shadow price of this hypothetical asset, in Appendix A we show that its valuation takes the familiar asset pricing form:

$$\theta_{it} = \mathbb{E}_{it} \left[ \beta \frac{u_c(c_{it+1}, \ell_{it+1})}{u_c(c_{it}, \ell_{it})} (y_{it+1} + \theta_{it+1}) \right]. \quad (6)$$

Thus, human wealth is the expected value of the stochastically discounted sum of next period's earnings plus the continuation value of human capital.

A complication that arises in this setting is that the current valuation of an asset depends on its effect on a person's marital bargaining power, if and when that person gets married in the future. If buying an asset does not increase the individual's utility once married, that asset would be worth less to them than otherwise. In this paper we do not attempt to explicitly estimate the effect on marital bargaining power of owning more shares of a hypothetical asset.<sup>4</sup> Rather, we make the simplifying assumption that bargaining between newly married couples can be represented through the symmetric Nash bargaining solution. In other words, the ex-post Pareto weights of spouses do adjust in response to pre-marital investments, and they do so through the effect of pre-marital investments on the outside options of spouses and on the marital surplus. In Appendix A we show that this assumption implies that any effect on human capital valuations operates exclusively through a single person's continuation value in marriage.

Another complication relates to how the hypothetical asset is allocated upon divorce. We assume that, in such circumstances, sole ownership of the asset based on individual  $i$ 's labor income would go to person  $i$ , and that other assets, possibly including a claim on alimony, would be allocated to the ex-spouse as compensation. The reason we assume that  $i$  takes ownership of the hypothetical asset is that we are valuing  $i$ 's human capital, which they would own upon divorce as well.<sup>5</sup> This assumption, along with the one described in the previous paragraph, allows us to derive tractable formulas for valuing one's own human capital. We exploit this tractability in the empirical analysis.

---

<sup>4</sup>In fact, this is a very interesting question in its own right but would require a much more sophisticated approach to modeling household interactions.

<sup>5</sup>One can, of course, be ordered to pay alimony out of human capital returns in the real world. However, alimony is usually a fixed amount of money, so changes in earnings affect the earners' net-income, not their spouses. Thus, alimony is better represented as an extra allocation of financial assets to the ex-spouse than an allocation of human capital, which is how we model it.

**The Role of Borrowing Constraints.** As discussed in Appendix A, we follow the ‘non-traded asset’ approach originally suggested by Lucas Jr (1978) to value human capital. This approach implicitly assumes that agents may trade human capital at the margin. This does not mean that agents do not face borrowing constraints in their portfolio of real assets  $\mathbf{a}_{it}$  (or  $\mathbf{a}_{(ij)t}$  if married). Rather, it means that even agents who are borrowing constrained in terms of different financial assets may contemplate selling a marginal unit of human capital at the appropriate price. There is an important and intuitive reason for this feature of the model. Our exercise recovers a price at which agents would *choose* not to trade away their human capital. For individuals that are borrowing constrained this price will clearly tend to be lower than for similar, but unconstrained, individuals. This is because constrained individuals would benefit from selling human capital as this would move them away from their borrowing constraint. Therefore this lower price is exactly what we want to recover because borrowing constrained individuals have lower valuations of their future earnings than unconstrained individuals. Indeed, future earnings are worth less to individuals who cannot access them in advance, and the way we have structured our exercise allows us to explicitly quantify this effect.

### 3 Estimating Human Wealth

Our approach to estimation of human wealth features two sequential steps. In the first step we apply the methods developed in Escanciano, Hoderlein, Lewbel *et al.* (2016) to recover nonparametric estimates of marginal utility functions, and of the deterministic component of the discount factor ( $\beta$ ). These are then used in a second step to obtain nonparametric estimates of human wealth. We overview both steps in detail, even though only the identification results for the second step are novel. The careful description of the first step greatly aids in understanding our identification and estimation approach in the second step.<sup>6</sup>

---

<sup>6</sup>We also outline a new procedure to accommodate the use of biennial data in the first step.

## 3.1 Identification

### 3.1.1 Nonparametric Marginal Utility Function Identification

It is helpful to use compact notation  $q = (c, \ell)$  and  $q' = (c', \ell')$  to represent the current and future choices of an arbitrary individual. The consumption decision of an individual who is not at a corner solution is described by the following intertemporal Euler equation:

$$u_c(q) = \beta \mathbb{E} [u_c(q')R' | q]. \quad (7)$$

This condition is written for the return  $R'$  on an arbitrary asset;  $R'$  could be the return on any asset traded by a subset of agents. Conditioning on current choices  $q$  is equivalent to conditioning on the entire information set because all relevant information is acted upon and reflected in these decisions.<sup>7</sup>

We begin by rewriting equation (7) in a form that replaces the expectation operator with the associated integral over the space of  $q'$ . In this integral the future marginal utilities are weighted by a factor corresponding to the product of (i) the conditional expectation of future rates of return and (ii) the Markov (transition) kernel estimator describing transitions from  $q$  to  $q'$ . The notation we use for this weighting factor is  $\psi(q, q') = \mathbb{E} [R' | q, q'] \times f(q' | q)$ , where  $f$  is the conditional density of  $q'$ . The Euler equation (7) can be represented as

$$u_c(q) - \beta \int u_c(q') \psi(q, q') dq' = 0. \quad (8)$$

As explained by Escanciano et al., this is a homogeneous Fredholm integral equation of the second kind. The solution for  $u_c(q)$  given  $\beta$  is well known. However, in our case both  $u_c(q)$

---

<sup>7</sup>From the point of view of an econometrician, the right-hand-side of Euler equation (7) is a function of consumption and leisure choices that depend on the (yet unknown) future state of the world  $\Omega'$ . That is, one could write (7) as

$$u_c(q(\Omega)) = \beta \mathbb{E} [u_c(q'(\Omega'))R'(\Omega') | \Omega].$$

For notational simplicity we omit the  $(\Omega, \Omega')$ .

and  $\beta$  must be determined, which leads to a question of identification.

**Finite Support Case.** Identification is easiest to understand if we restrict ourselves to the case in which the space of  $q_i$  is a finite number  $M$  of consumption/leisure pairs. Formally, the support is  $q \in \{q^1, q^2, \dots, q^M\}$ . Under this assumption we can rewrite the Euler equation (8) at any current choice vector  $q^k$  as

$$u_c(q^k) - \beta \sum_{m=1}^M u_c(q^m) \psi_d(q^k, q^m) = 0, \quad (9)$$

where  $\psi_d$  is a discrete analogue of the transition function  $\psi$ . Rather than solving a complicated integral equation, identification in this finite example requires solving a linear system. Writing equation (9) in matrix notation, this entails solving

$$(I - \beta\Psi)U_c = 0, \quad (10)$$

where  $U_c = (u_c(q^1), u_c(q^2), \dots, u_c(q^M))'$ , and  $\Psi$  is a  $M \times M$  matrix, with  $\Psi_{km} = \psi_d(q^k, q^m)$ .

This system has a nontrivial solution with  $U_c \gg 0$  only if  $\det(I - \beta\Psi) = 0$ , which is true if  $\beta^{-1}$  is an eigenvalue of  $\Psi$ . In such cases the solution for  $U_c$  will depend on the eigenvector of  $\Psi$  associated with the eigenvalue  $\beta^{-1}$ . Thus,  $\beta$  is identified as the inverse of any eigenvalue of  $\Psi$  such that  $\beta \in (0, 1)$ , and  $U_c$  is identified as the solution of the homogeneous system for the associated eigenvector. In general,  $\Psi$  may have multiple eigenvalues larger than unity, thus only set identification is achieved in the finite support case. It is worth noting that  $\Psi$  is not simply a transition matrix (whose largest eigenvalue would be 1), but rather a transition matrix multiplied (element-wise) by expected asset returns  $\mathbb{E}[R'|q, q']$ .

**General Case.** Proof of identification in the general case where  $q$  has a continuous support requires functional analysis, but is reminiscent of the finite support case above. One first

defines a linear operator  $A$  that, when applied to the unknown function  $u_c(q)$ , results in

$$(Au_c)(q) = \beta \int u_c(q') \psi(q, q') dq'. \quad (11)$$

By definition this implies that  $u_c = \beta Au_c$ . In the case that  $u_c$  and  $Au_c$  are positive valued (marginal utility is positive) and  $A$  is a compact operator, a solution for  $u_c$  exists only if  $\beta = 1/\rho(A)$ , where  $\rho(A)$  is the largest eigenvalue (spectral radius) of the operator  $A$ .<sup>8</sup> Therefore, if these assumptions are maintained, a unique value of  $\beta$  and a unique function  $u_c$  solve equation (8) and point identification is achieved.

### 3.1.2 Nonparametric Human Wealth Identification

We now turn to the second step and to the question of nonparametric identification of  $\theta_{it}$  in equation (6). Relying on the results derived above, we posit that  $\beta$  and the marginal utility function are identified.

Next, we introduce the vector  $\mathbf{z}$  containing variables that summarize an individual's information set. Unlike the estimation of the marginal utility function, we now also consider individuals who may be credit constrained. Therefore current consumption and leisure may not fully summarize each individual's information set, posing a problem when approximating expectations. For this reason we make the following assumption:

**Definition (Conditional Equivalence of Expectations):** *Expectations are conditionally equivalent with respect to the vector  $\mathbf{z}$  if for any individual  $i$  and time period  $t$*

$$\mathbb{E}_{it} \left[ \beta \frac{u_c(q')}{u_c(q)} (y' + \theta') \right] = \mathbb{E} \left[ \beta \frac{u_c(q')}{u_c(q)} (y' + \theta') \mid \mathbf{z} = \mathbf{z}_{it} \right].$$

Conditional equivalence of expectations holds if  $\mathbf{z}_{it}$  is sufficient to span the current information

---

<sup>8</sup>In the infinite dimensional case a linear compact positive operator has one positive eigenvector and its corresponding eigenvalue is equal to the spectral radius of the operator. Hence, we have uniqueness in this case.

set of any individual  $i$  and time  $t$ .<sup>9</sup> Assuming this is the case, we can rewrite the human wealth valuation equation (6) with  $\theta_{it}$  replaced by a function  $\theta(j, \mathbf{z})$ , where  $j$  is the age of the individual:

$$\theta(j, \mathbf{z}) = \mathbb{E} \left[ \beta \frac{u_c(q')}{u_c(q)} (y' + \theta(j+1, \mathbf{z}')) \middle| \mathbf{z} \right]. \quad (12)$$

This is a functional equation, similar to the Euler equation analyzed above.

We rewrite equation (12) as an integral equation after operating two substitutions. First, define  $\delta(j, \mathbf{z}, \mathbf{z}') = \mathbb{E}[\beta(u'_c/u_c)|j, \mathbf{z}, \mathbf{z}'] \times f_{\mathbf{z}'|Z}^j(\mathbf{z}'|j, \mathbf{z})$ , where  $f_{\mathbf{z}'|Z}^j$  is the age-specific conditional density of  $\mathbf{z}'$ . Each  $\delta(j, \mathbf{z}, \mathbf{z}')$  can be described as an appropriately discounted density function for  $\mathbf{z}'$  at age  $j$ , for given conditioning set  $\mathbf{z}$ . Second, we define  $g(j, \mathbf{z}) = \mathbb{E}[\beta(u'_c/u_c)y'|j, \mathbf{z}]$ , which subsumes the expected discounted value of the human wealth dividend. It follows that the human wealth equation can be written as

$$\theta(j, \mathbf{z}) = g(j, \mathbf{z}) + \int \theta(j+1, \mathbf{z}') \delta(j, \mathbf{z}, \mathbf{z}') d\mathbf{z}'. \quad (13)$$

Comparing the above functional equation to the integral form of the intertemporal Euler equation, the key difference is that eq. (13) is an *inhomogeneous* Fredholm integral equation of the second kind. The lack of homogeneity is due to the presence of the term  $g(j, \mathbf{z})$ , which introduces the age-dependent intercept in equation (13).

One can provide conditions for a unique solution of equation (13) by exploiting the deterministic nature of age transitions. We begin by defining the vector-valued functions

$$\begin{aligned} \Theta(\mathbf{z}) &= (\theta(1, \mathbf{z}), \theta(2, \mathbf{z}), \dots, \theta(J-1, \mathbf{z}), \theta(J, \mathbf{z}))' \\ G(\mathbf{z}) &= (g(1, \mathbf{z}), g(2, \mathbf{z}), \dots, g(J-1, \mathbf{z}), 0)', \end{aligned}$$

where  $J$  is an arbitrarily old age at which earnings are zero. Furthermore, arrange the age-

---

<sup>9</sup>This assumption can be verified by testing whether ex-post measures of realized shocks are in fact orthogonal to observed consumption expenditures.

specific transition functions into a  $J \times J$  matrix

$$\Delta(\mathbf{z}, \mathbf{z}') = \begin{pmatrix} 0 & \delta(1, \mathbf{z}, \mathbf{z}') & 0 & \dots & 0 \\ 0 & 0 & \delta(2, \mathbf{z}, \mathbf{z}') & & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & & \delta(J-1, \mathbf{z}, \mathbf{z}') \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \quad (14)$$

This matrix conforms with  $\Theta(\mathbf{z}')$  in a way that permits the following representation of the integral equation (13):

$$\Theta(\mathbf{z}) = G(\mathbf{z}) + \int \Delta(\mathbf{z}, \mathbf{z}')\Theta(\mathbf{z}')d\mathbf{z}'. \quad (15)$$

Like in Escanciano et al., we next define a linear operator  $B$  composed of a finite set of age-specific linear operators  $B_j$ . Each age-specific operator satisfies

$$(B_j\theta)(j+1, \mathbf{z}) = \int \delta(j, \mathbf{z}, \mathbf{z}')\theta(j+1, \mathbf{z}')d\mathbf{z}'. \quad (16)$$

Then, the operator  $B$  is defined as follows:

$$B = \begin{pmatrix} 0 & B_1 & 0 & \dots & 0 \\ 0 & 0 & B_2 & & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & & B_{J-1} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \quad (17)$$

This ensures that  $B$  is a linear operator such that:

$$(B\Theta)(\mathbf{z}) = \int \Delta(\mathbf{z}, \mathbf{z}')\Theta(\mathbf{z}')d\mathbf{z}'. \quad (18)$$

Using this definition within equation (15), the function  $\Theta$  is uniquely determined to be  $\Theta = (I - B)^{-1}G$ , provided the operator  $I - B$  has a well defined inverse.<sup>10</sup> The invertibility of  $I - B$  depends on the properties of  $B$ : this follows from the assumption that, for a large enough age  $J$ , the value of human wealth is zero, which leads to  $B$  being upper triangular and hollow. The simple intuition for this identification result becomes apparent if one solves the pricing equation (15) recursively, starting from the last age in which human wealth has a non-zero value. That is, in the last period one can use the fact that the human wealth value next period is zero to solve for the value of human wealth in the current period, which will simply be  $g(J - 1, z)$ . This solution can then be stored and used to solve for the value of human wealth one period prior, allowing for a backward recursion up to the initial age.

**Finite example.** When the support of  $\mathbf{z}$  is restricted to be finite, so that  $\mathbf{z} \in \{\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^M\}$ , proof of a unique solution for  $\Theta$  amounts to proving a unique solution for a linear system. In such a case each operator  $\delta(j, \mathbf{z}, \mathbf{z}')$  becomes a matrix. Each such matrix is an element of the block matrix  $\Delta$ , which is hollow and upper triangular. Applying this to our human wealth equation we have  $\Theta = G + \Delta\Theta$ , the solution of which is  $\Theta = (I - \Delta)^{-1}G$ , if the inverse exists. Because  $\Delta$  is hollow and upper triangular, all eigenvalues of  $(I - \Delta)$  are unity, and therefore the inverse exists and  $\Theta$  has a unique solution.

---

<sup>10</sup>This is the case if the operator  $B$  is compact and it does not have an eigenvalue of exactly unity.

## 3.2 Empirical Implementation

We consider a sample  $\{q_{it}, q'_{it}, z_{it}, z'_{it}, R'_{it}, y'_{it}, j_{it}, j'_{it}\}$ . Index  $i \in N$  denotes an element within the set  $N$  of observed individuals. Index  $t \in \tau(i)$  identifies the periods within the set of years  $\tau(i)$  for which the variables are observed for individual  $i$ . Crucially, the set  $\tau(i)$  includes sample years for which  $i$  is observed in both the current and subsequent sample periods: that is, both choices  $q_{it}$  and  $q'_{it}$  must be observed. For example, if a person is observed for three subsequent waves of the data panel that person contributes two observations to the sample.

We let  $\tau_o(i) \subset \tau(i)$  be the subset of observations in which individual  $i$  is at an interior solution for assets (that is, not borrowing constrained). Then, we proceed sequentially. First, we describe our implementation of the estimator of the marginal utility function by Escanciano et al. This description features our extension to allow a mixture of annual and biennial observations. Second, we overview our estimator of the human wealth valuation functions.

### 3.2.1 Estimation of the Marginal Utility Function

The first step in the estimation of the marginal utility function is to replace the linear operator  $A$  in equation (11) with the estimator

$$(\hat{A}u_c)(q) = \sum_{i=1}^N \sum_{t \in \tau_o(i)} u_c(q'_{it}) R'_{it} \phi_{it}(q). \quad (19)$$

The weighting functions  $\phi_{it}(q)$  deliver the locally weighted average estimator (Nadaraya-Watson estimator) of the conditional expectation in equation (11). Mechanically, we construct the weighting functions as

$$\phi_{it}(q) = \frac{K_{it}(q)}{\hat{f}(q)}, \quad (20)$$

where

$$\hat{f}(q) = \sum_{i=1}^N \sum_{t \in \tau_o(i)} K_{it}(q), \quad (21)$$

and

$$K_{it}(q) = K^H(q - q_{it}). \quad (22)$$

The function  $K^H(\cdot)$  is a multivariate Gaussian kernel function with bandwidth vector  $H$ .

Since the estimator  $\hat{A}$  has a finite dimensional range (unlike the true  $A$ ),  $\hat{A}$  has a finite number of eigenvalues and eigenfunctions, which can be computed by solving a linear system. Hence any eigenfunction  $\hat{u}_c(q)$  of  $\hat{A}$  must be a linear combination of the functions  $\phi_{it}(q)$ , i.e.  $\hat{u}_c(q) = \sum_{i=1}^N \sum_{t \in \tau_o(i)} b_{it} \phi_{it}(q)$  for some set of coefficients  $b_{it}$ . Using this result, the empirical counterpart of the intertemporal Euler equation can be re-written as

$$\sum_{i=1}^N \sum_{t \in \tau_o(i)} b_{it} \phi_{it}(q) = \hat{\beta} \sum_{i=1}^N \sum_{t \in \tau_o(i)} \left[ \sum_{m=1}^N \sum_{s \in \tau_o(s)} b_{ms} \phi_{ms}(q'_{it}) \right] R'_{it} \phi_{it}(q). \quad (23)$$

The left side of the equation above simply replaces  $u_c(q)$  with its estimator. The right side first uses equation (19) to replace the expectation in the Euler equation (7) with its estimator, and then also replaces  $u_c(q'_{it})$  with its estimator (the part in square brackets). Straightforward algebra shows that a sufficient condition for the Euler equation above to have a solution is

$$b_{it} - \hat{\beta} \sum_{m=1}^N \sum_{s \in \tau_o(s)} b_{ms} \phi_{ms}(q'_{it}) R'_{it} = 0, \quad (24)$$

for every  $i \in N$  and  $t \in \tau_o(i)$ . This can be rewritten in matrix form with  $\Phi$  being a square matrix with elements  $\Phi_{kl} = \phi_l(q'_k) R'_k$ , and  $\mathbf{b}$  being a vector containing the coefficients  $b_{it}$  appropriately concatenated. Thus the restrictions in equation (24) are

$$(I - \hat{\beta} \Phi) \mathbf{b} = \vec{0}. \quad (25)$$

Letting  $\lambda^*$  be the largest eigenvalue of  $\Phi$  in absolute value, and  $b^*$  the associated eigenvector,

the estimators of  $\beta$  and  $u_c(q)$  are respectively

$$\hat{\beta} = \frac{1}{|\lambda^*|} \quad (26)$$

$$\hat{u}_c(q) = \sum_{i=1}^N \sum_{t \in \tau_o(i)} b_{it}^* \phi_{it}(q). \quad (27)$$

With no loss of generality  $\hat{u}_c(q)$  can be scaled to have a unit norm.

**Incorporating Biennial Data.** If data are available only at two year intervals, the empirical counterpart of the Euler equation becomes

$$\hat{u}_c(q) = \hat{\beta}^2 \sum_{i=1}^N \sum_{t \in \tau_o(i)} \hat{u}_c(q'_{it}) R'_{it} \phi_{it}(q), \quad (28)$$

where  $q'_{it}$  denotes decisions taken two years after  $t$ , and  $R'_{it}$  is the return on assets over a two-year period. One complication of adding biennial data is that the interpretation of the largest eigenvalue of  $\Phi$  is  $1/\beta^2$  for this set of observations and  $1/\beta$  for the annual observations. Estimating separately in the two samples, or transforming the annual data into a biennial panel, implies a reduction in sample size and loss of precision. Thus, we seek an estimation approach that allows for the joint use of annual and biennial data.

Our solution entails the transformation  $\hat{\beta}^2 = \hat{\beta} \beta_0$ , where  $\beta_0$  is some initial estimate of  $\hat{\beta}$  (possibly based only on the annual data sample). Then, after replacing  $R'_{it}$  with  $\tilde{R}'_{it} = \beta_0 R'_{it}$ , we employ biennial observations in the following moment condition:

$$\hat{u}_c(q) = \hat{\beta} \sum_{i=1}^N \sum_{t \in \tau_o(i)} \hat{u}_c(q'_{it}) \tilde{R}'_{it} \phi_{it}(q). \quad (29)$$

Now the largest eigenvalue of  $\Phi$  can be correctly interpreted as  $1/\beta$  for all observations in the extended sample. However, the estimates of  $\hat{\beta}$  and  $\hat{u}_c$  are conditional on  $\beta_0$ , hence they can

be improved upon if a better estimate of  $\beta_0$  becomes available. We replace  $\beta_0$  by  $\hat{\beta}$  and re-estimate, iterating this procedure until  $\hat{\beta}$  is approximately equal to the guess  $\beta_0$  and no further improvement is feasible.

### 3.2.2 Estimation of Human Wealth

Point estimates of marginal utility can be recovered for each individual choice observed in our sample by evaluating the function  $\hat{u}_c(q)$  at  $q_{it}$ .<sup>11</sup> Given these point estimates, the next step is to construct an estimator for the age-specific human wealth valuation functions. For ease of presentation, we describe this estimator for the case of annual observations. Later we illustrate how to extend it to biennial observations.

We begin by estimating the value of the expected dividend function  $g(j, \mathbf{z})$  in equation (13) using the Nadaraya-Watson estimator:

$$\hat{g}(j, \mathbf{z}) = \sum_{i=1}^N \sum_{t \in \tau_j(i)} \hat{\beta} \frac{\hat{u}_c(q'_{it})}{\hat{u}_c(q_{it})} y'_{it} \gamma_{it}(\mathbf{z}). \quad (30)$$

To this purpose we need to link individuals of the same age across time periods. First, we define  $\tau_j(i)$  as the singleton set of years in which individual  $i$  is  $j$  years old. Then, given an individual  $i$  of age  $j$  in period  $t$ , we let  $\tau_{j(it)}(m)$  be the singleton set of years in which an individual  $m$  is exactly the same age as individual  $i$  in period  $t$  (that is, exactly  $j$  years old). The weighting functions  $\gamma_{it}(\mathbf{z})$  are then constructed as

$$\gamma_{it}(\mathbf{z}) = \frac{K_{it}^z(\mathbf{z})}{\sum_{m=1}^N \sum_{t \in \tau_{j(it)}(m)} K_{mt}^z(\mathbf{z})}, \quad (31)$$

where  $K_{it}^z(\mathbf{z})$  is a multivariate kernel function.<sup>12</sup>

<sup>11</sup>This can be done even if the particular person-year observation  $(i, t)$  refers to a credit constrained individual, hence not used in the estimation procedure described above.

<sup>12</sup>Here we follow Li and Racine (2007) by defining  $\mathbf{z}^c$  and  $\mathbf{z}^d$  to be the sub-vectors of continuous and discrete variables contained in  $\mathbf{z}$ . The multivariate kernel function for a given  $\mathbf{z}_{it}$  can then be written as  $K_{it}^z(\mathbf{z}) =$

Next, we form estimators of the  $\theta(j, \mathbf{z})$  functions. We re-write equation (13) replacing all functions by their estimators:

$$\hat{\theta}(j, \mathbf{z}) = \hat{g}(j, \mathbf{z}) + \sum_{i=1}^N \sum_{t \in \tau_j(i)} \hat{\theta}(j+1, \mathbf{z}'_{it}) \hat{\beta} \frac{\hat{u}_c(q'_{it})}{\hat{u}_c(q_{it})} \gamma_{it}(\mathbf{z}). \quad (32)$$

Because we have an estimate of  $\hat{g}(j, \mathbf{z})$ , the only obstacle to obtain an estimate of the current human wealth function  $\hat{\theta}(j, \mathbf{z})$  is that the future function  $\hat{\theta}(j+1, \mathbf{z}')$  is so far unknown. However, as it is clear from equation (32), the entire function  $\hat{\theta}(j+1, \mathbf{z}')$  need not be known. Rather, one only needs to estimate its value at the subset of points  $\mathbf{z}'_{it}$  that are observed (the data points for which  $t \in \tau_j(i)$ ). Stacking all  $\hat{\theta}(j+1, \mathbf{z}'_{it})$  into vectors  $\tilde{\Theta}_{j+1}$ , and similarly stacking the  $\hat{g}(j+1, \mathbf{z}'_{it})$  into vectors  $\tilde{G}_{j+1}$ , we can re-write equation (32) in compact form as  $\tilde{\Theta}_j = \tilde{G}_j + \Gamma_j \tilde{\Theta}_{j+1}$ . The matrix  $\Gamma_j$  has number of rows equal to the number of observations stacked in  $\tilde{\Theta}_j$  and number of columns equal to the number of observations stacked in  $\tilde{\Theta}_{j+1}$ . The elements of  $\Gamma_j$  are

$$[\Gamma_j]_{mi} = \hat{\beta} \frac{\hat{u}_c(q'_{it})}{\hat{u}_c(q_{it})} \gamma_{it}(z_{mt}). \quad (33)$$

Each column of  $\Gamma_j$  includes the transition kernel and stochastic discount factor of a given individual  $i$ . For each such individual  $i$  there is a corresponding age  $j+1$  human wealth estimate contained in  $\tilde{\Theta}_{j+1}$ . However, each row of  $\Gamma_j$  is evaluated at the data vector  $z_{mt}$  of a (usually) different individual  $m$ .<sup>13</sup> For each such individual  $m$  there is a corresponding age  $j$  human wealth estimate contained in  $\tilde{\Theta}_j$ . If the data sample is unbalanced one may have

---

$(\prod_{z_s \in \mathbf{z}^c} K^{h_s}(z_s - z_{s,it})) \times (\prod_{z_s \in \mathbf{z}^d} \mathbf{1}_{\{z_s = z_{s,it}\}})$ . The first product includes univariate gaussian kernel functions with bandwidth  $h_s$ . The second product includes indicator functions, which ensure the kernel has positive value for an observation with the corresponding values of all discrete variables, and zero otherwise. For example, this means that female data will have no influence on the conditional expectation for a male observation, and vice-versa.

<sup>13</sup>We structure our data so that an observation consists of pairs  $\{\mathbf{z}_{it}, \mathbf{z}'_{it}\}$  and  $\{j_{it}, j'_{it}\}$ . If an individual in the sample is observed over multiple periods, then this individual contributes multiple observations to the estimation. Therefore, while a kernel  $\gamma_{it}(\mathbf{z})$  is generally evaluated using data from different individuals, it may also be evaluated (at most once) at different data points for the same individual. This effectively treats the evaluation of one's own future realizations, where the individual is one year older, as a different observation.

different numbers of observations at different ages. In this case  $\Gamma_j$  will not be square and the lengths of  $\tilde{\Theta}_j$  and  $\tilde{\Theta}_{j+1}$  will differ.

We combine vectors  $\tilde{\Theta}_j$  and  $\tilde{G}_j$  into larger vectors  $\tilde{\Theta} = (\tilde{\Theta}'_1, \dots, \tilde{\Theta}'_{J-1}, \tilde{\Theta}'_J)'$  and  $\tilde{G} = (\tilde{G}'_1, \dots, \tilde{G}'_{J-1}, \vec{0}')$ , where  $J$  is an arbitrarily old age by which all individuals have either died or retired. We then also arrange all matrices  $\Gamma_j$  into a block matrix  $\Gamma$ ,

$$\Gamma = \begin{pmatrix} 0 & \Gamma_1 & 0 & \dots & 0 \\ 0 & 0 & \Gamma_2 & \dots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \Gamma_{J-1} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \quad (34)$$

Using this notation the set of  $j$ -specific equations  $\tilde{\Theta}_j = \tilde{G}_j + \Gamma_j \tilde{\Theta}_{j+1}$  can be written compactly as:

$$\tilde{\Theta} = \tilde{G} + \Gamma \tilde{\Theta}. \quad (35)$$

Because  $(I - \Gamma)$  is invertible<sup>14</sup>, one can directly solve for  $\tilde{\Theta} = (I - \Gamma)^{-1} \tilde{G}$ .

Finally, to obtain estimators of the complete functions  $\theta(j, \mathbf{z})$ , rather than just estimating at a subset of observed data points, we return to equation (32). Because the point estimates of  $\hat{\theta}(j+1, \mathbf{z}'_{it})$  are now available (they are the elements of  $\tilde{\Theta}$ ), equation (32) can be evaluated at *any* point  $\mathbf{z}$ . Thus, the vector of estimators for the age-specific human wealth valuation functions  $(\hat{\theta}(1, \mathbf{z}), \hat{\theta}(2, \mathbf{z}), \dots, \hat{\theta}(J, \mathbf{z}))'$  has now been obtained.

**Incorporating Biennial Data.** We now consider the case in which some sample data are at annual frequency while others are only available at a biennial frequency. To accommodate

---

<sup>14</sup>Note that  $\Gamma$  is upper triangular and hollow, therefore  $(I - \Gamma)$  is upper-triangular with ones on the leading diagonal, so its determinant is also one.

this discrepancy we denote  $\tau_j^1(i)$  and  $\tau_j^2(i)$  as the set of annual and biennial sample years, respectively, in which  $i$  was of age  $j$ . For an observation drawn during a period of biennial sampling, equation (32) can be rewritten by iterating the valuation equation one-year further into the future:

$$\hat{\theta}(j, \mathbf{z}) = \hat{g}^1(j, \mathbf{z}) + \hat{g}^2(j, \mathbf{z}) + \sum_{i=1}^N \sum_{t \in \tau_j^2(i)} \hat{\theta}(j+2, \mathbf{z}'_{it}) \frac{\hat{u}_c(q'_{it})}{\hat{u}_c(q_{it})} \gamma_{it}(\mathbf{z}). \quad (36)$$

It is understood that, for the biennial sample,  $q'_{it}$  and  $\mathbf{z}'_{it}$  are data observations two years into the future. The functions  $\hat{g}^1(j, z)$  and  $\hat{g}^2(j, z)$  are estimates of the conditional expectation of discounted earnings, one and two years ahead, for a  $j$  year old individual with current state vector  $\mathbf{z}$ . These estimates are computed as in equation (30), where  $\hat{g}^1$  is estimated using data from the annual sample period and  $\hat{g}^2$  is estimated using data from the biennial sample period.

As before, we form vectors  $\tilde{\Theta}$  and  $\tilde{G}$ , as well as a matrix  $\Gamma$  such that  $\tilde{\Theta} = \tilde{G} + \Gamma\tilde{\Theta}$ . Some elements of  $\tilde{\Theta}$  and  $\tilde{G}$  are based on annual observations using equation (32) with  $\tau(i)$  replaced by  $\tau^1(i)$ , and others are based on biennial observations using equation (36).

Matrix  $\Gamma$  is somewhat more complicated because rows corresponding to biennial observations must conform with columns of  $\tilde{\Theta}$  corresponding to values two years ahead. Thus,  $\Gamma$  now must have the form

$$\Gamma = \begin{pmatrix} 0 & \Gamma_1^1 & \Gamma_1^2 & 0 & \dots & 0 \\ 0 & 0 & \Gamma_2^1 & \Gamma_2^2 & \dots & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & 0 & \Gamma_{J-1}^1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (37)$$

where  $\Gamma_j^1$  and  $\Gamma_j^2$  are constructed as explained in equation (33). The reason we now have

two blocks in each row of  $\Gamma$  is to allow rows corresponding to annual observations to multiply  $\tilde{\Theta}_{j+1}$ , and rows corresponding to biennial observations to multiply  $\tilde{\Theta}_{j+2}$ . Rows of  $\Gamma_j^1$  corresponding to annual observations will contain elements as in equation (33), whereas rows corresponding to biennial observations will consist of zeros. Conversely, rows of  $\Gamma_j^2$  corresponding to annual observations will contain all zeros, but rows corresponding to biennial observations will contain elements as in equation (33). After constructing such a matrix  $\Gamma$  we can solve for  $\tilde{\Theta} = (I - \Gamma)^{-1}\tilde{G}$  as before.

The last step is to construct an estimator for the general function  $\hat{\theta}(j, \mathbf{z})$ , once estimates have been recovered by computing  $\tilde{\Theta}$  at the observed sample points. This requires a weighting of equations (32) and (36). We define numbers of annual and biennial observations  $n^1 = \sum_{i=1}^N \sum_{t \in \tau^1(i)} 1$  and  $n^2 = \sum_{i=1}^N \sum_{t \in \tau^2(i)} 1$ . Using these counts we form the estimator as

$$\begin{aligned} \hat{\theta}(j, \mathbf{z}) = & \hat{g}^1(j, \mathbf{z}) + \frac{n^1}{n^1 + n^2} \left( \hat{\beta} \sum_{i=1}^N \sum_{t \in \tau_j^1(i)} \hat{\theta}(j+1, \mathbf{z}'_{it}) \frac{\hat{u}_c(q'_{it+1})}{\hat{u}_c(q_{it})} \gamma_{it}(\mathbf{z}) \right) \\ & + \frac{n^2}{n^1 + n^2} \left( \hat{g}^2(j, \mathbf{z}) + \hat{\beta}^2 \sum_{i=1}^N \sum_{t \in \tau_j^2(i)} \hat{\theta}(j+2, \mathbf{z}'_{it}) \frac{\hat{u}_c(q'_{it})}{\hat{u}_c(q_{it})} \gamma_{it}(\mathbf{z}) \right). \end{aligned} \quad (38)$$

Weighting in this way ensures that, if there are only a small number of biennial observations, these observations have a limited influence on the estimated functions.

### 3.3 Data

To obtain an empirical counterpart of the estimator in equation (23), and to recover the marginal utilities, we need panel data on consumption and leisure, as well as historical asset returns and proxies for information available to individuals when making decisions. The sample must include observations recorded over a sufficiently long time interval so to identify the aggregate risk component of the transition kernels.

Hence the basic data requirement for the estimation of marginal utilities, discount factor and human wealth values is a sample  $\{q_{it}, q'_{it}, \mathbf{z}_{it}, \mathbf{z}'_{it}, R'_{it}, y'_{it}, j_{it}, j'_{it}\}_{i \in N, t \in \tau(i)}$ , where each vector  $q$  denotes a pair of consumption and leisure choices; the vector  $\mathbf{z}$  includes variables that approximate the information set of the decision makers;  $R$  is a historical real return from deferred consumption; and  $j$  denotes age.

It turns out that the Panel Study of Income Dynamics contains much of what we need. We use panel data from the PSID spanning the years 1967-2016. We assume that repeated observations for the same individuals in this data set satisfy the required weak dependence condition.

Construction of  $q_i$  and  $q'_i$  involves collecting earnings and consumption data. Labour earnings is always observed. However, a fairly complete set of consumption expenditures is observed at the household level only after 1997. Before that date only selected categories of consumption were recorded regularly.<sup>15</sup> For this reason we build on the approach of Attanasio and Pistaferri (2014) to approximate household consumption expenditure in periods when information is incomplete. This method relies on the ever larger availability of consumption expenditures in the PSID post-1997. The procedure effectively estimates a demand system to impute consumption to PSID families observed in years before 1997. There are five advantages to this approach: (i) it relies on information from a single data set, making variable linkages straightforward; (ii) one can test how closely trends in consumption inequality are replicated by the imputation procedure using within-sample verification for the period during which complete expenditure data are available; (iii) since the PSID stretches all the way back to the late 1960s, this procedure delivers the longest consumption panel database currently available for the US; (iv) average consumption per household can be scaled to replicate its historical evolution; (v) last but not least, expenditure categories in the PSID appear to match NIPA counterparts reasonably well.

---

<sup>15</sup>If one goes back all the way to 1967, only food expenditures were regularly measured.

As a proxy for real asset returns we set  $R'_i$  to be the one-year treasury constant maturity rate minus realized annual CPI inflation, when using annual data prior to 1997. As the survey becomes biannual after 1997, we switch to the two-year treasury constant maturity rate minus realized CPI inflation.<sup>16</sup>

Finally, to obtain an empirical counterpart of the human wealth estimator in equation (32) we use a set of conditioning variables that approximate the information set available to agents. The vector  $\mathbf{z}_i$  contains (i) observable individual characteristics, such as gender, education, industry, occupation, marital status, number of dependent children, and (ii) unobserved type, as we discuss below.

### 3.4 Estimation of Unobserved Types

The data vector  $\mathbf{z}_i$  includes an unobserved type,  $\eta_i$ , which we allow to vary along two dimensions of heterogeneity. The first dimension captures differences in life-cycle earning profiles, identified from variation in the growth rates of earnings. The second dimension subsumes unobserved differences in wealth, which we measure by gauging the dispersion of consumption growth rates over the life cycle of different sample members.

To estimate heterogeneous types we use an approach similar to that of Bonhomme, Lamadon, and Manresa (2017). That is, we employ a k-medians grouping algorithm to separate life-cycle averages of ‘informative’ variables into clusters, where cluster membership is a type. Cluster membership is then represented through categorical variables. The idea is that variation in income and consumption growth paths conveys information about, respectively, permanent heterogeneity in income and in idiosyncratic access to wealth used to smooth consumption.

To test whether our grouping procedure does a good job of estimating unobserved heterogeneity, and to establish the number of types used to model each dimension of heterogeneity, we follow the reasoning of Cunha, Heckman, and Navarro (2005). These authors suggest that,

---

<sup>16</sup>These time series are publicly available from FRED. We also experiment with real returns for other assets.

if agents know their own wealth and earnings type, they should act upon such information and make choices that are consistent with their type. More generally, it should be possible to identify heterogeneity due to ex-ante types because individuals respond to this information and act on it. If unobserved types' heterogeneity contributes to idiosyncratic earnings growth and life-cycle consumption changes, then the latter should be helpful in predicting observable long term choices.

Following Cunha et al., we illustrate this point using the decision to attend college. Let  $S_i$  denote the college decision of individual  $i$ , taking value one if the individual completes college and zero otherwise. To the extent that heterogeneity  $\eta_i$  affects earnings growth, one would expect that  $E(S_i|\eta_i) \neq 0$ . Given the relationship between unobserved types and economic outcomes (such as earnings and consumption), schooling choices should be related with the (ex-post) level of earnings growth, or with the idiosyncratic dispersion of consumption growth rates. By the same token, if one could control directly for the underlying type  $\eta_i$ , the expectation of college completion should no longer respond to these observable measures of ex-post earnings or consumption. This line of reasoning offers a natural way to test whether our grouping procedure identifies the relevant "type" variation.

If the grouping algorithm successfully captures the relevant heterogeneity, the type indicator should crowd out the statistical effect of earnings profiles (and, similarly, of consumption dispersion) on college status. We find that allowing for three types to represent earnings heterogeneity is sufficient to remove any direct effect of earnings growth on the expectation of college completion. In the case of wealth heterogeneity we only need two types for the conditional expectation of college completion to be independent of consumption growth dispersion. Having established the cardinality of the types' sets, we also corroborate our clustering by verifying that adding further types does not result in significant drops in within-type variances.

### 3.5 Human Wealth in the Survey of Consumer Finances

The PSID provides a long panel data set with sufficient information to carry out our human wealth estimation exercise. For a wider analysis of inequality, however, we rely on the Survey of Consumer Finances (SCF), which provides much more detail on both the value and composition of households' wealth, going back to 1989. By design, the SCF also captures the upper tail of the wealth distribution far better than the PSID.

While the SCF is valuable for the measurement of wealth portfolios, we cannot estimate human wealth valuations from it because of its lack of a panel dimension. However, since we have recovered the entire function  $\Theta(\mathbf{z})$ , we can directly evaluate equation (32) at *any* data point  $\mathbf{z}$ . For this reason we are able to obtain point estimates of human wealth in any data set where an appropriate counterpart of the vector  $\mathbf{z}$  is available. Unfortunately not every variable in the data vector  $\mathbf{z}$  is observed in the SCF. In particular, unobserved types  $\eta$  cannot be estimated from repeated cross-sections of data. In addition, some variables are only observed for the household head (for example education attainment and age). To deal with this problem we impute the *full distribution* of the missing variables estimated from the PSID. It is important to impute the distributions of missing variables, rather than use their conditional expectations, because the latter would average out heterogeneity and lead to underestimates of inequality.

**Set Imputation.** We perform this 'set imputation' by first partitioning the data vector  $\mathbf{Z}$  into observed variables  $\mathbf{Z}^+$  and unobserved variables  $\mathbf{Z}^-$ . We then define the conditional distribution function  $\Pi(\mathbf{Z}^-|\mathbf{Z}^+)$ . Because  $\mathbf{Z}^-$  takes discretely many values, this distribution can be viewed as a probability mass function  $\Pi(\mathbf{Z}^-|\mathbf{Z}^+) = \{\pi_1(\mathbf{Z}^+), \pi_2(\mathbf{Z}^+), \dots, \pi_M(\mathbf{Z}^+)\}$ , where  $M$  is the number of points in the support set of  $\mathbf{Z}^-$ . In turn, each  $\pi_m(\mathbf{Z}^+)$  can be estimated using the Nadaraya-Watson kernel estimator using PSID data.

Next, we expand the SCF data set so that it replicates the cross-sectional variation of  $\mathbf{Z}^-$ . We do this by creating  $M$  versions of the extended SCF sample, one for each of the

$M$  points in the support set of  $\mathbf{Z}^-$ . Hence, each such version imputes a different point in the support of  $\mathbf{Z}^-$ . The sample weight for observation  $i$  in version  $m \in M$  of the data is rescaled by  $\pi_m(\mathbf{Z}_i^+)$ . Finally, we stack these subsamples into a single data set. Each original SCF observation appears  $M$  times in the expanded data set, but the total weight of these  $M$  replications is rescaled to equal the sample weight of the original (individual) observation. Human wealth can then be computed for each observation in the new expanded sample, and analysis can proceed by using the adjusted sample weights.

## 4 Estimation Results

Using the methods described above we obtain a set of estimates for the overall lifetime wealth of households in the SCF sample, in conjunction with a detailed decomposition of the relative composition of each household's wealth portfolio at a point in time.

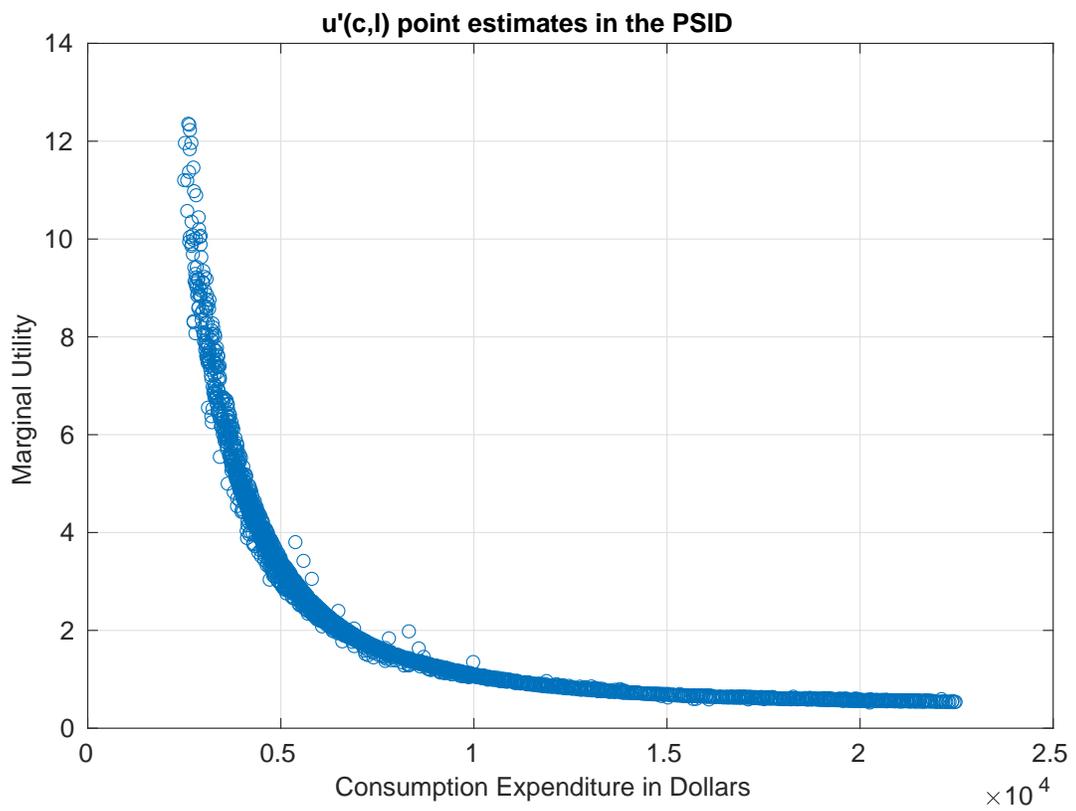
### 4.1 Estimates of Marginal Utility and Human Wealth

Non-parametric estimates of the marginal utility of consumption are plotted in Figure 1. Consistent with theory, marginal utility is highly non-linear at low expenditure levels and flattens out at high expenditures.

**Human wealth estimates: PSID sample.** One can use the estimated marginal utility function and the human wealth valuation function (in equation 38) to stochastically discount earnings and assign a pecuniary value to the human capital held by different households in the PSID sample.

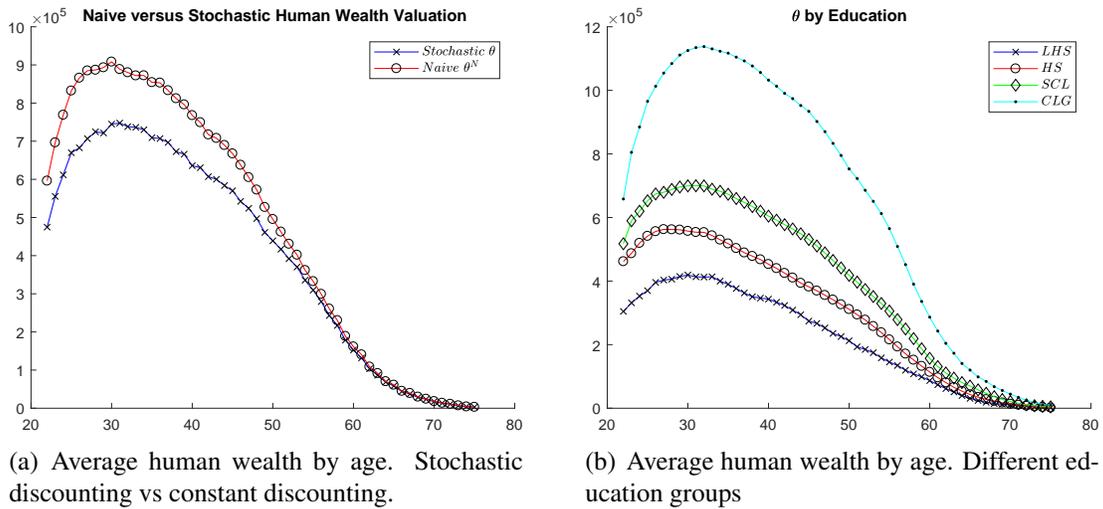
The left panel of Figure 2 plots the average value of human wealth at different ages, and contrasts it to the value estimated using a constant discount factor. The constant discount factor is set equal to the average of all realizations of the stochastic discount factor used in estimating the marginal utility function, and hence is equivalent to the theoretical price of a

Figure 1: Marginal utility as a function of consumption expenditures.



one-period risk free bond. The figure indicates that fixed discounting results in a significant overestimation of human wealth, with the largest discrepancy largest around the time when human wealth peaks. This confirms simulation-based results in Huggett and Kaplan (2016).

Figure 2: Average human wealth over the life cycle. Values in 2016 dollars.



Human wealth exhibits a hump-shape, with a steep drop after age 50 as retirement approaches. Under state-dependent discounting the average value peaks at around \$800K. This average is based on a wide sample of individuals, including some who do not work. Non-employment risk is explicitly accounted for in our estimation, which considers periods of null earnings as one of the possible outcomes of a worker's job employment history. The right panel of Figure 2 plots the value of human wealth by education group. As one might expect, there are large differences in both scale and shape. At the peak, college graduates hold more than twice as much human wealth as high-school graduates, and more than three times as much human wealth as high-school drop-outs. Human wealth differentials become progressively smaller as retirement gets closer and the pecuniary value of human capital converges towards very low values.

Two interesting observations can be made at this stage. First, younger households hold most of their wealth in a very illiquid asset, which exposes them to significant risks. For ex-

ample, health shocks might affect their labor supply and reduce the present value of human capital. Second, the early peak in human wealth also suggests that the direct (that is, unmediated by assets) contribution of human wealth to overall inequality must occur at relatively younger ages, when human wealth still accounts for a high proportion of wealth portfolios. We revisit some of these issues in the context of our SCF sample.

Figure 3: Estimation error due to constant discounting of human capital, by education group.

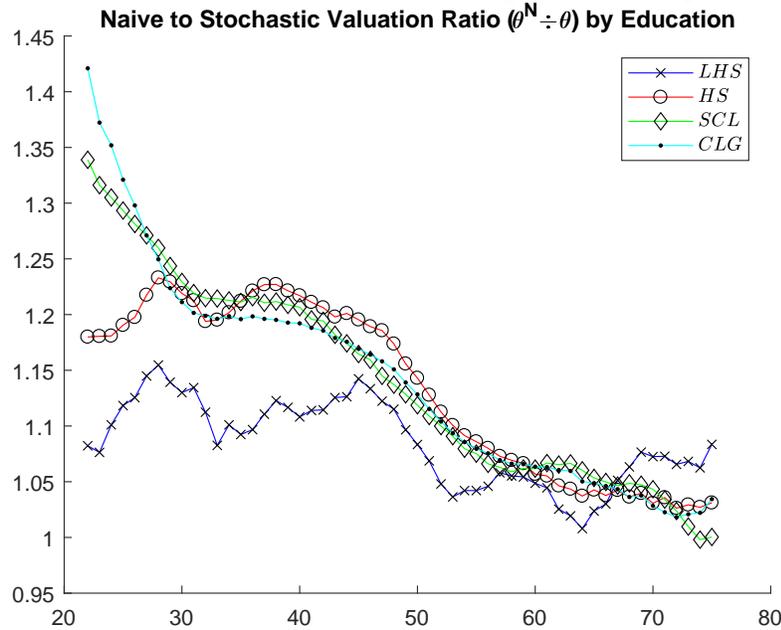
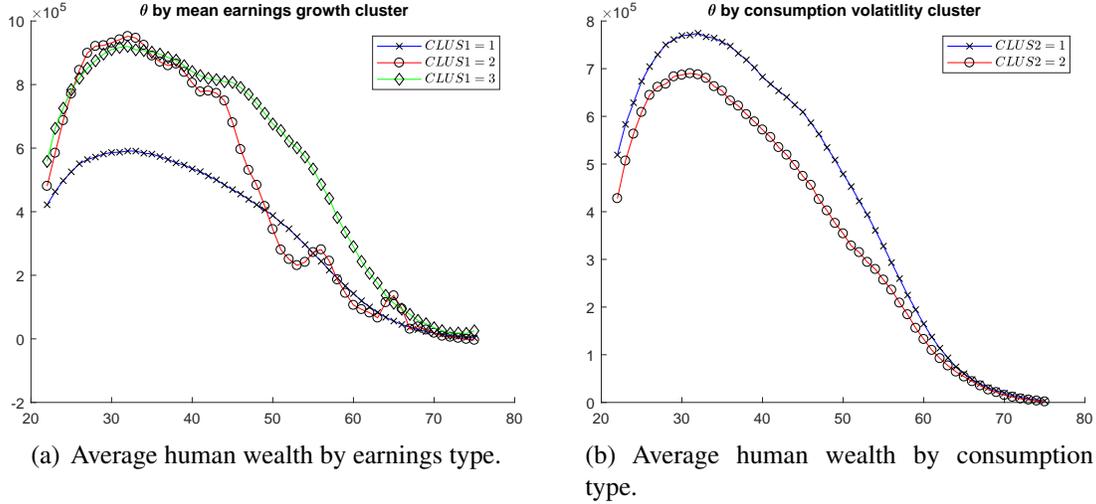


Figure (3) reports the ratio of our estimates under fixed and flexible discounting for different education groups. The approximation error is much larger for higher education levels, especially at younger ages when several educated households may suffer for liquidity constraints. Finally, in Figure 4 we report the evolution of average human wealth over the life cycle for different latent type clusters. The left panel in Figure 4 reports the human wealth for the three earnings' types, while the right panel reports estimates for the two clusters identified using consumption dispersion. Differences are fairly large and persistent. Unobserved heterogeneity identified through earnings growth induces a near doubling of human wealth at the peak. This gap is remarkable because it suggests the presence of a strong latent source of

within group wealth heterogeneity among observationally similar households.

Figure 4: Average human wealth over the life cycle by latent type. Values in 2016 dollars.



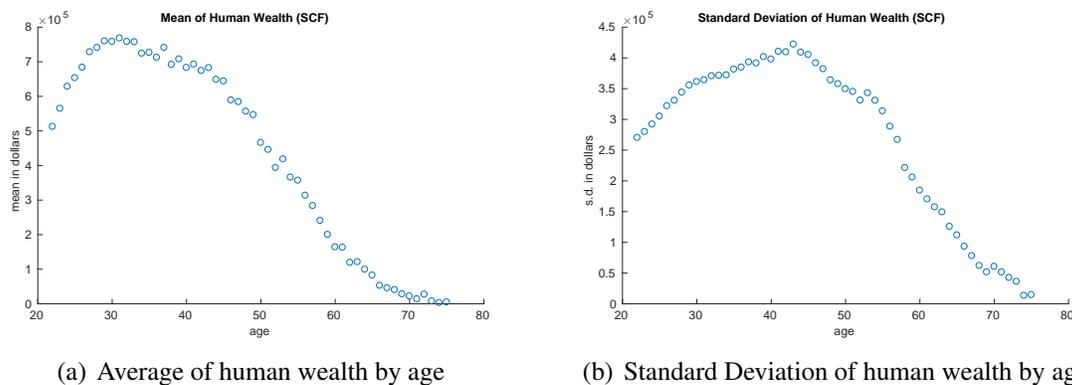
**Human wealth estimates: SCF sample.** Using the valuation function in equation (38) we recover an estimate of human wealth for each household in the SCF sample.<sup>17</sup> This allows us to quantify the relative size of human wealth in their wealth portfolio. In the left panel of Figure 5 we report the life-cycle evolution of average human wealth. Both shape and scale of average human wealth closely track those estimated from PSID data and plotted in Figure 2.

The right panel of Figure 5 also plots the standard deviation of human wealth, which is roughly half the size of the average human wealth at any given age. For example, average human wealth peaks at just below \$800,000 (per household), when the standard deviation stands at roughly \$350,000. This means that two standard deviations below the average corresponds to a value close to zero, while adding two standard deviations doubles the average value. Interestingly, dispersion remains fairly high until age 50. Hence, the contribution of human wealth to overall inequality is largest between ages 35 and 55. The fact that dispersion remains elevated long after average human wealth has started its decline indicates that some workers are

<sup>17</sup>As we discussed, this requires linking the full distribution of unobservable types to each observation in the SCF.

exiting full time employment relatively young: low employment and non-employment risks are explicitly accounted for by our estimation, which considers periods of null or low earnings as possible outcomes of each worker’s history.

Figure 5: Mean and standard deviation of human wealth over the life cycle (SCF sample, values in 2016 dollars).



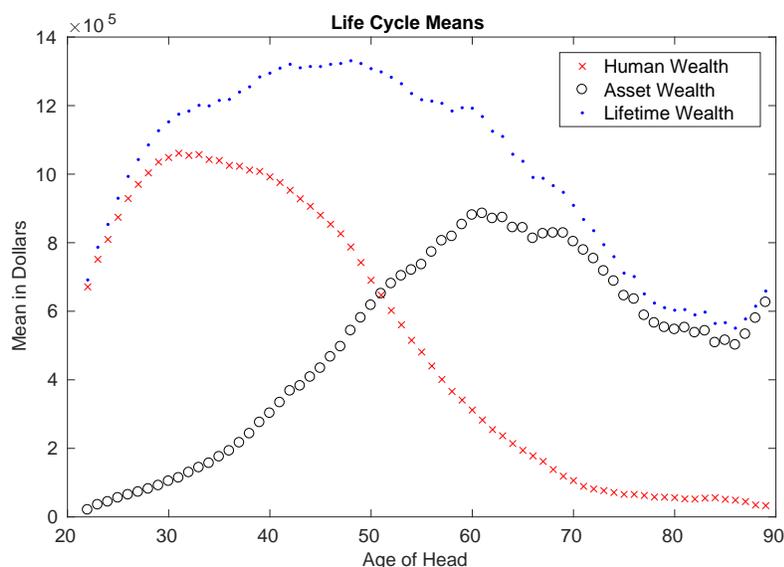
## 4.2 The Relative Smoothness of Lifetime Wealth

Figure 6 reports the average value of human and non-human wealth components, as well as of total wealth, for all households in our sample.<sup>18</sup> It is apparent that total wealth is remarkably stable over the life cycle, and certainly more so than its individual components. Young households’ portfolios are heavily skewed towards human wealth, which makes shocks impacting labor supply or health very costly for them. In fact, any shocks are likely to be poorly insured among young adults because their net worth (non-human wealth) tends to be low.

The value of human wealth peaks early in life, around age 30. This is well before the peak age for earnings and draws attention to two key aspects: first, the expected length of remaining working life is important when putting a price tag on a stream of labor earnings; second, earlier investments in human capital carry a higher return while its depreciation may

<sup>18</sup>One could easily disaggregate non human wealth in its different components.

Figure 6: Average human wealth, net worth and total wealth (the sum of human wealth and net worth) over the lifecycle. Values in 2016 dollars.

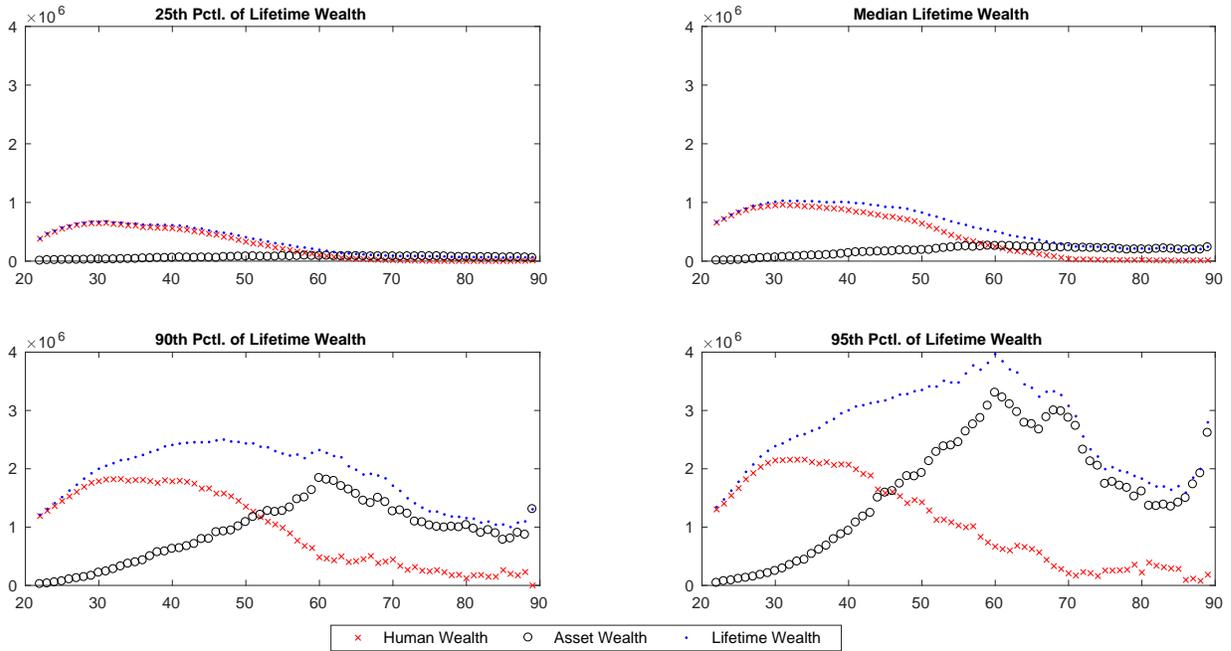


become more severe with age. These observations suggest that using current earnings as a measure of cross-sectional inequality is problematic, something that we revisit below.

**The changing composition of wealth over the life-cycle.** The contrast between human and non-human wealth is striking. Assets net worth peaks around age 60 and effectively accounts for all wealth after age 70. Yet, net worth accounts for a relatively small fraction of total wealth until age 40. Given these patterns, total wealth peaks early (around age 30) and, while declining to roughly 1/3 of its peak value by age 80, it exhibits less extreme proportional variation than its individual components over the course of the life cycle. This relative ‘smoothness’ of total wealth over the life cycle is consistent with the finding that a large chunk of total wealth is determined early in life in the form of human wealth. Then, over time, total wealth changes shape, shifting from illiquid human wealth to more liquid net worth. In this sense, the process of aging mostly changes the composition of wealth, while its total value varies less. It is also interesting that the peak in total wealth occurs later in life for people who already own some assets when young. This will become apparent below, where we examine the evolution of life

cycle wealth for different percentiles of the wealth distribution.

Figure 7: Average human wealth, net worth and total wealth over the lifecycle, by percentile of total wealth. Values in 2016 dollars.



**Wealth over the life cycle: rich and poor households.** Figure 5 documents significant dispersion in the distribution of human wealth at any given age. This dispersion captures permanent differences in the value of discounted earnings. To gauge possible differences in the evolution of wealth holdings over the life cycle we contrast wealth patterns for households at the 25<sup>th</sup>, 50<sup>th</sup>, 90<sup>th</sup> and 95<sup>th</sup> percentile of total wealth at each given age. The results are plotted in Figure 7 and confirm the presence of significant heterogeneity in wealth portfolios over the life cycle.

Households at the lower end of the total wealth distribution hold little asset wealth at any age, while richer households exhibit larger net worth at relatively early ages. Interestingly, we observe that human wealth plays a quantitatively large role even at the top end of the wealth distribution, representing a significant share of the aggregate at early ages. As we

anticipated, total wealth peaks early among poorer households. These households do not appear to convert their human wealth in equivalent amounts of net worth later in life. In turn this is due to the fact that households with relatively lower earnings often behave as hand-to-mouth consumers so that their lifetime wealth profile is almost always downward sloping and significantly less smooth. The message does not change at all if we condition on the percentiles of each individual variable separately, instead of conditioning on percentiles of total wealth, as shown in the Appendix Figure 11. This confirms that the ranking of total wealth broadly lines up with the rankings of human wealth early in life and with the ranking of asset wealth at later ages.

Results disaggregated by percentile also offer two useful insights into the fanning out of wealth inequality over the life cycle. First, inequality is mostly due to human wealth in the first twenty years of working life. Second, because inequality in human wealth is smaller than inequality in asset wealth at almost all ages, the relative share of asset wealth in households' portfolios must play a key role in the evolution of permanent income.

## **5 Wealth and Permanent Income Concentration over Time**

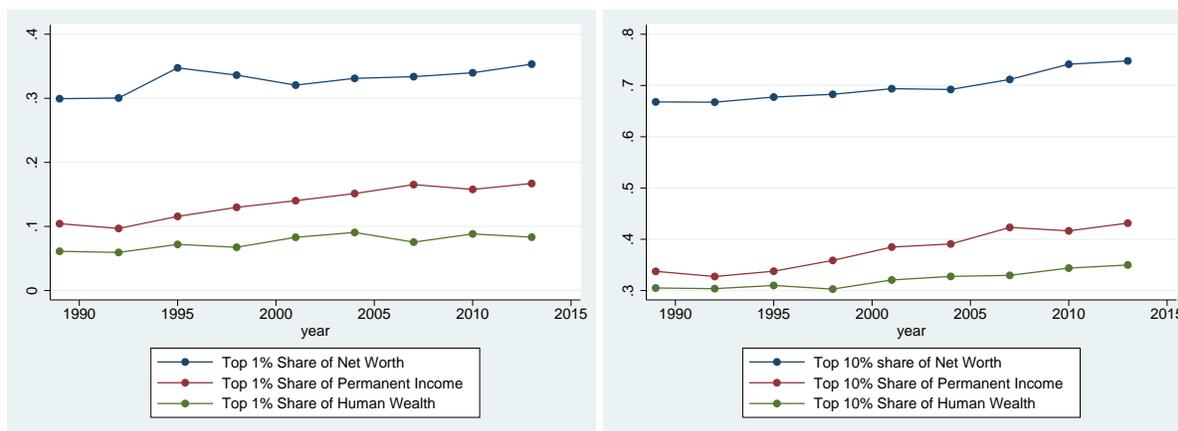
We next use our extended SCF data to measure concentration of net worth, as in Bricker, Henriques, Krimmel *et al.* (2016), and we contrast it to measures of concentration for human wealth, total wealth and permanent income. To facilitate comparison with existing studies, and to provide a simple summary of changes in inequality over time, we begin by plotting measures of wealth concentrations for different variables in different years.

### **5.1 Permanent Income vs Asset Wealth**

In the two panels of Figure 8 we report the share of, respectively, net worth, permanent income and human wealth held by the top 1% (left panel) and the top 10% (right panel) of households

in the distributions of the respective variables. Of course, the households at the top of each distribution may be different.

Figure 8: Concentration of net worth (assets), permanent income, and human wealth.



(a) Shares held by households in the top 1% of the respective distribution

(b) Shares held by households in the top 10% of the respective distribution

Accounting for human wealth fundamentally changes our view of inequality and its evolution. First, permanent income is much less concentrated than assets (real/financial wealth). The share of permanent income held by the top 1% is on average 15% of the total. This is roughly half of the share of net worth held by the top 1%, which is well over 30% of total assets. This suggests that lifetime wealth and permanent income are significantly *less concentrated* than net worth. Similar patterns can be observed when looking at concentration measures for the richest top 10% of households. The share of permanent income held by the richest 10% of households is roughly half the share of assets wealth held by households at the top of the net worth distribution.

However, Figure 8 also shows that the growth of permanent income concentration has been strong over the past 35 years. Both human wealth and net worth have become significantly more concentrated between 1989 and 2016, but the increase for permanent income concentration has far outpaced that for net worth. This is remarkable for three reasons: first, the share of permanent income is almost twice as large in 2016 as it was in 1989; second, in the face of

the heated debate on wealth inequality, since 1989 the speed at which permanent income has concentrated in the hands of the richest households is almost twice as large as the well-known increase in the share of real/financial wealth; third, human wealth concentration is unlikely to account for the speedy growth in permanent income.

**Comparing measures of concentration.** For a more nuanced view of the changing concentration of economic resources, Tables 1 and 2 report shares of different variables held by the top 10% of households. Table 1 reports the share of each variable held by the top 10% of households in the distribution of *that variable*, while Table 2 reports the share of each variable held by the top 10% of households in the distribution of net worth (real/financial assets).

year	Net Worth	Human Wealth	Lifetime Wealth	Earnings	Permanent Income
	(1)	(2)	(3)	(4)	(5)
1989	0.668	0.369	0.392	0.386	0.414
1992	0.667	0.371	0.387	0.392	0.408
1995	0.677	0.374	0.400	0.403	0.426
1998	0.683	0.367	0.415	0.377	0.434
2001	0.693	0.371	0.437	0.413	0.468
2004	0.691	0.379	0.445	0.417	0.481
2007	0.712	0.381	0.469	0.439	0.506
2010	0.741	0.387	0.462	0.460	0.513
2013	0.747	0.400	0.478	0.460	0.528
2016	0.768	0.404	0.519	0.499	0.569

Table 1: This table reports the share of variable “X” in the hands of the households in the top 10% of the distribution of that same variable “X”. For example, the share of earnings held by the households in the top 10% of the distribution of earnings.

Human wealth exhibits significantly lower concentration at the top than earnings. This suggests that a significant component of earnings concentration is due to transitory shocks. In fact, human wealth has the lowest concentration among all variables, as permanent heterogeneity is mitigated by relatively short working lives and the presence of idiosyncratic shocks

<b>year</b>	<b>Net Worth</b>	<b>Human Wealth</b>	<b>Lifetime Wealth</b>	<b>Earnings</b>	<b>Permanent Income</b>
	(1)	(2)	(3)	(4)	(5)
1989	0.668	0.120	0.320	0.224	0.368
1992	0.667	0.115	0.299	0.222	0.353
1995	0.677	0.122	0.318	0.239	0.369
1998	0.683	0.122	0.348	0.238	0.390
2001	0.693	0.125	0.388	0.279	0.436
2004	0.691	0.118	0.396	0.272	0.453
2007	0.712	0.116	0.424	0.291	0.474
2010	0.741	0.115	0.420	0.305	0.487
2013	0.747	0.120	0.429	0.305	0.495
2016	0.768	0.118	0.487	0.339	0.550

Table 2: This table reports the share of variable “X” in the hands of the households in the top 10% of the distribution of Net Worth. For example, the share of earnings held by the households in the top 10% of the distribution of net worth.

that may depreciate human capital and limit the extent of excess-returns from skills (see for example Gallipoli and Turner, 2011).

Looking at trends, concentration has risen for both earnings and human wealth. However, the top 10% share of earnings appears to have grown more, confirming that the relative dispersion of transitory shocks has become larger (Gottschalk, Moffitt, Katz *et al.*, 1994; Heathcote, Perri, and Violante, 2010). Restricting attention only to households at the top of the distribution of asset wealth (Table 2), only earnings exhibit increasing concentration while human wealth concentration is effectively unchanged. This suggests that the share of human wealth held by asset-rich households has not grown over the past decades, and that the growth in their earnings’ share is largely due to low-persistence income shocks. Hence, increasing wealth concentration in the hands of asset-rich households does not appear to derive from better labor market returns.

These observations illustrate why looking at earning flows (rather than human wealth stocks) would be misleading, whether studying levels or trends. They also reveal useful in-

formation about the composition of households at the top of different wealth distributions. Human wealth has become more concentrated (Table 1), yet the share of human wealth belonging to households at the top of the net worth distribution has not changed (Table 2). Thus, large increases in the top share of permanent income imply that the contribution of net worth to household portfolios has risen significantly over time. Put simply, asset-rich households have not become more likely to be at the top of the human wealth distribution. rather, human wealth has, over time, become a less important determinant of inequality in permanent income. Being rich in human wealth is less important for permanent income in 2016 than it was in 1989.

<b>year</b>	<b>Net Worth</b>	<b>Human Wealth</b>	<b>Lifetime Wealth</b>	<b>Earnings</b>	<b>Permanent Income</b>
	(1)	(2)	(3)	(4)	(5)
1989	0.296	0.057	0.117	0.104	0.149
1992	0.299	0.057	0.110	0.104	0.147
1995	0.347	0.059	0.133	0.118	0.173
1998	0.336	0.057	0.145	0.113	0.172
2001	0.319	0.059	0.158	0.152	0.189
2004	0.330	0.059	0.169	0.136	0.200
2007	0.333	0.060	0.181	0.157	0.219
2010	0.339	0.061	0.172	0.156	0.228
2013	0.352	0.063	0.181	0.151	0.243
2016	0.384	0.067	0.225	0.199	0.266

Table 3: This table reports the share of variable “X” in the hands of the households in the top 1% of the distribution of that same variable “X”. For example, the share of earnings held by the households in the top 1% of the distribution of earnings.

Tables 3 and 4 report similar information for the top 1% of households. Our findings broadly confirm those for the top 10% of households. The top shares of all variables are rising; yet, there is no evidence that asset-rich households are holding a larger share of human wealth. The fact that their share of human wealth does not rise, while their share of permanent income increases significantly, indicates the growing role of real/financial wealth as a share of

<b>year</b>	<b>Net Worth</b>	<b>Human Wealth</b>	<b>Lifetime Wealth</b>	<b>Earnings</b>	<b>Permanent Income</b>
	(1)	(2)	(3)	(4)	(5)
1989	0.296	0.010	0.115	0.043	0.146
1992	0.299	0.012	0.108	0.048	0.143
1995	0.347	0.013	0.131	0.061	0.165
1998	0.336	0.015	0.144	0.061	0.166
2001	0.319	0.015	0.156	0.071	0.176
2004	0.330	0.015	0.168	0.078	0.186
2007	0.333	0.015	0.180	0.077	0.204
2010	0.339	0.013	0.172	0.067	0.203
2013	0.352	0.013	0.180	0.069	0.236
2016	0.384	0.017	0.225	0.114	0.256

Table 4: This table reports the share of variable “X” in the hands of the households in the top 1% of the distribution of net worth. For example, the share of earnings held by the households in the top 1% of the distribution of net worth.

permanent income. Permanent income inequality appears to be the byproduct of the changing composition of wealth, rather than of increasing dispersion in human wealth.

A caveat is in order. All these results are static and one must consider the possibility that rising human wealth concentration early in life generates rising asset wealth inequality later in life. In the following section we revisit some of these issues.

**Lifetime consumption, post-tax human wealth and human endowment values.** In Table 5 we explore three alternative ways to gauge changes in the concentration of resources and contrast them to previous results about the concentration of human wealth and permanent income. For comparison, columns (1) and (2) in Table 5 reproduce results from Table 1 above.

In our baseline estimation we assume that human wealth is a function of the possible paths of life-cycle labor supply observed in the data. This approach implicitly values realized, rather than potential, earnings. To account for the value of the total endowment of human wealth (including the opportunity cost of time) we re-estimate human wealth under the assumption that

every available hour is valued at its market value. This assumes out the effect of differences in lifetime labor supply. Column (3) in Table 5 shows that the resulting variable, denoted ‘human endowment value’, exhibits slightly lower concentration than our baseline human wealth measure. This is not unexpected: our human wealth estimates account for differences in the concentration of hours worked while the human endowment value ignores variation in hours worked. Nonetheless, the change in concentration over time closely tracks baseline estimates for human wealth, suggesting that wages rather than hours worked are responsible for the growing concentration of human capital.

In column (4) we report estimates of the concentration of post-tax human wealth. These estimates are obtained from the distribution of earnings after taxes and transfers.<sup>19</sup> This adjusted measure of human wealth appears only slightly less concentrated than our baseline human wealth estimates. Moreover it exhibits an almost identical pattern over time.

In the last column of the same table we also report a proxy of permanent income concentration based on lifetime consumption expenditures. This measure is obtained from the present value of life-cycle consumption and is estimated exactly like the human wealth value, computing the value of the flow of consumption expenditures rather than earnings. In this way we are able to contrast our measures of lifetime wealth to the distribution of resources as reflected in expenditures on non-durable consumption. Lifetime non-durable consumption appears less concentrated in levels than permanent income, and exhibits only a small increase between 1989 and 2016. While we know that the mapping from expendable income to actual consumption is mediated by a variety of taxes and formal and informal transfers, the cross-sectional distribution of lifetime consumption across households is also less concentrated than that of post-tax human wealth. This suggests that households only consume a fixed amount of their permanent income. Moreover, the amount of non-durable consumption accounts for a progressively smaller share for households at the top of the wealth distribution. Of course,

---

<sup>19</sup>To approximate post-tax earnings we use a power function adjustment whose properties are described in Guner, Kaygusuz, and Ventura (2014).

our measure of consumption is far from perfect and conveys little information about durable expenditures. Hence, we consider these results as a lower bound of the concentration of total expenditures. Even so, the difference in magnitudes is significant enough to suggest the presence of: (i) informal insurance channels that induce a significant amount of redistribution; and (ii) motives that induce households to save large amounts of their lifetime wealth. In fact, the discrepancy between permanent income and consumption concentration clearly indicates that, even at older ages, a large share of resources is being saved.

year	Human Wealth	Perm. Income	Human Endowm. Value	Human Wealth (post tax and transfers)	Lifetime Cons.
	(1)	(2)	(3)	(4)	(5)
1989	0.369	0.414	0.345	0.360	0.300
1992	0.371	0.408	0.351	0.362	0.300
1995	0.374	0.426	0.353	0.365	0.302
1998	0.367	0.434	0.346	0.358	0.296
2001	0.371	0.468	0.352	0.361	0.298
2004	0.379	0.481	0.358	0.369	0.304
2007	0.381	0.506	0.360	0.372	0.304
2010	0.387	0.513	0.365	0.378	0.307
2013	0.400	0.528	0.376	0.391	0.314
2016	0.404	0.569	0.384	0.395	0.319

Table 5: This table reports the share of variable “X” in the hands of the households in the top 10% of the distribution of that same variable “X”. For example, the share of earnings held by the households in the top 10% of the distribution of earnings.

## 5.2 The Mechanics of Increasing Inequality

The preceding analysis provides a portrayal of the historical patterns of US wealth concentration over the past few decades. Next we explore some related questions about the specific mechanisms at work. Has the composition of households’ wealth changed over the past 30 years? How do young and old households in 2016 compare to their counterparts in 1989? Has

demographic change contributed to growth in permanent income inequality? And which other factors played a role?

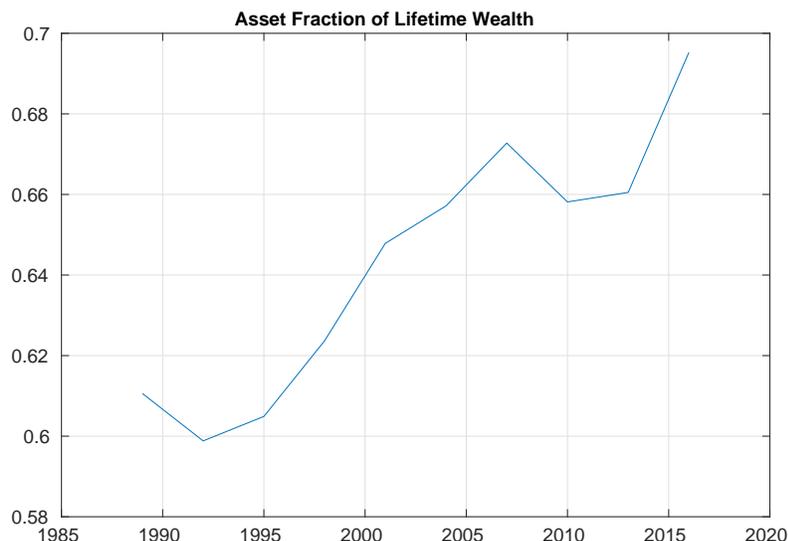
While it is clear that households in the top of the net worth distribution have been steadily increasing their share of assets, an offsetting effect might come from rising concentration of human wealth in the hands of a different set of households at the top of the human wealth distribution. In contrast, having the same subset of households sit at the top of both distributions would compound and exacerbate the concentration of permanent income. Hence, a key question is whether the joint probability of being near the top of both the human and asset wealth distributions has changed over time.

To answer these questions we perform several checks: (i) first, we directly measure how much of the stock of total wealth in different years is accounted for by asset wealth; (ii) we characterize the role of a changing age composition over the past few decades; (iii) we provide a way to account for the joint evolution of the distribution of asset wealth and human wealth.

**The changing importance of assets in households' portfolios.** So far we have provided indirect evidence that the relative size of asset wealth as a share of total wealth may have increased in the recent past. This has wide ranging implications for the nature and extent of wealth inequality. A larger share of liquid assets would imply a better ability on the side of households to respond to shocks like disability, unemployment and displacement; however, a diminished role for human wealth may indicate that asset accumulation is driven by factors other than hard work and higher early-life earnings. To assess the relative importance of asset wealth we use our SCF extended sample and calculate, for each year in the sample, the ratio of average asset wealth to total wealth across all households. The time series of this ratio is plotted in Figure 9. It is apparent that asset wealth has become progressively more important in households' portfolios. Starting from a value of around 60% in the early 1990s, asset wealth accounted in 2016 for almost 70% of total wealth. The only exceptions to this systematic growth pattern occurred during the recession of the late 2000s, when asset prices

and valuations were temporarily reduced. This phenomenon might partly due to the changing age composition in the United States. We consider this possibility below.

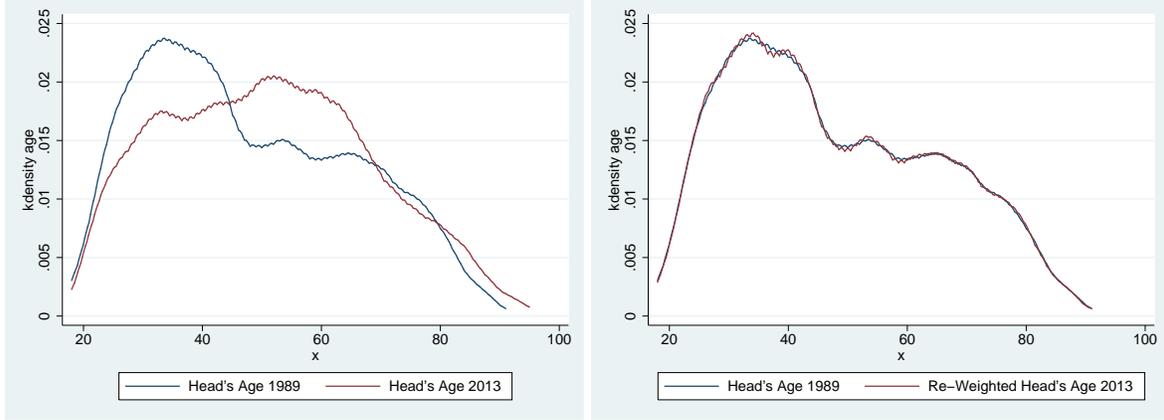
Figure 9: Asset wealth as a share of total household wealth (average by year).



**The role of demographic change.** One might hypothesize that the trend towards an older population is responsible for part of the dynamics of inequality. To examine this possibility we carry out a re-weighting exercise in the spirit of DiNardo, Fortin, and Lemieux (1996) that allows us to verify how inequality in the variables we observe would have changed had the age distribution stayed the same as in 1989. For each year we fit probit regressions with a full set of age dummy variables, then we use the predicted probabilities to transform the SCF sample weights into a new set of weights that forces the age distribution to be constant. Figure 10 illustrates how the age distribution changed between 1989 and 2016, and how our counterfactual weights reshape the 2016 age distribution.

Using the age correction described above we produce counterfactual versions of the concentration tables. Table 6 reproduces the top 10% shares of each variable under the counterfactual re-weighting. Interestingly this counterfactual analysis indicates almost no change in human wealth concentration, in contrast to the five percentage point increase we observe in

Figure 10: Actual age distributions in 1989 and 2016 (left) and re-weighted counterfactual age distribution in 2016 with actual age distribution in 1989 (right).



the unadjusted baseline analysis. This discrepancy illustrates that the rising concentration of human wealth follows from the fact that, in 2016, a much smaller segment of the population is at their peak of human wealth. Hence fewer households outside the top 10% have large human wealth stocks in 2016 than in 1989. This mechanically accounts for the run-up in human wealth concentration.

Finally, in our counterfactual age-adjusted analysis the concentration of lifetime wealth and permanent income grows higher but not as high as in the baseline. This implies that asset wealth has become increasingly important as a component of overall inequality.

**Who got richer? Some inequality accounting.** *Very Preliminary* In this section we address the question of which type of households accounts for the increasing concentration of permanent income (PI). By definition, the share of PI held by the households in the top 10% of the PI distribution in period  $t$  can be written as

$$s_{PI}^{10}(t) = \frac{PI^{10}(t)}{PI^{10}(t) + PI^{90}(t)},$$

where  $PI^x(t)$  is the aggregate value of permanent income held by households in the top  $x\%$  of the distribution of PI in year  $t$ . The share  $s_{PI}^{10}(t)$  can be split between the net worth component

<b>Year</b>	<b>Net Worth</b>	<b>Human Wealth</b>	<b>Earnings</b>	<b>Lifetime Wealth</b>	<b>Perm. Income</b>
	(1)	(2)	(3)	(4)	(5)
1989	0.668	0.305	0.408	0.337	0.360
1992	0.669	0.298	0.397	0.322	0.339
1995	0.683	0.303	0.404	0.335	0.355
1998	0.690	0.299	0.405	0.353	0.375
2001	0.706	0.313	0.424	0.378	0.407
2004	0.703	0.314	0.415	0.378	0.411
2007	0.720	0.316	0.437	0.404	0.441
2010	0.759	0.306	0.444	0.379	0.419
2013	0.764	0.307	0.441	0.393	0.433

Table 6: This table reports the counterfactual share of variable “X” in the hands of the households in the top 10% of the distribution of that same variable “X”. For example, the share of earnings held by the households in the top 10% of the distribution of earnings. These counterfactuals hold that age distribution constant as it was in 1989.

and the human wealth component as follows

$$s_{PI}^{10}(t) = \frac{NW^{10}(t) + HW^{10}(t)}{NW^{10}(t) + NW^{90}(t) + HW^{10}(t) + HW^{90}(t)}.$$

Here  $NW^n$  denotes the aggregate net worth value held by the top  $n\%$  of the PI marginal distribution and  $HW^n$  denotes the aggregate human wealth held by the same set of households.

We define the wedge  $\delta_x(t)$  as the value of variable  $x$  that, if redistributed from the top 10% to the bottom 90% of households, would make their relative share of  $x$  in year  $t$  identical to that observed in 1989. That is, we define  $\delta_x(t)$  as the value such that

$$\frac{x^{10}(t) - \delta_x(t)}{x^{90}(t) + \delta_x(t)} = \frac{x^{10}(1989)}{x^{90}(1989)}.$$

The wedge  $\delta_x(t)$  allows us to compute counterfactual inequality values for the distribution of

a variable  $x$ , which can be used to account for changes in the concentration of  $x$  over time.<sup>20</sup>

For instance, if the relative distribution of net worth had not changed between 1989 and year  $t$ , the counterfactual share of permanent income ( $PI$ ) held by households in the top 10% of the PI distribution in period  $t$  would be

$$\widetilde{s}_{PI}^{10}(t, \delta_{NW}) = \frac{NW^{10}(t) + HW^{10}(t) - \delta_{NW}(t)}{NW^{10}(t) + NW^{90}(t) + HW^{10}(t) + HW^{90}(t)},$$

an expression that features the net worth wedge  $\delta_{NW}(t)$  only at the numerator.

Similar reasoning suggests that, absent changes in the distribution of human wealth after 1989, the counterfactual share of permanent income held by the the top 10% of households in the PI distribution would be

$$\widetilde{s}_{PI}^{10}(t, \delta_{HW}) = \frac{NW^{10}(t) + HW^{10}(t) - \delta_{HW}(t)}{NW^{10}(t) + NW^{90}(t) + HW^{10}(t) + HW^{90}(t)}.$$

By definition, any change in the observed value of the share  $s_{PI}^{10}$  that is not accounted for by  $\delta_{NW}$  and  $\delta_{HW}$  must be due to changes in the relative importance of net worth and human capital in the composition of permanent income. Therefore the difference  $\Delta_{NW} = s_{PI}^{10}(t) - \widetilde{s}_{PI}^{10}(t, \delta_{NW})$  measures how much of  $s_{PI}^{10}(t)$  is due to increasing concentration of net worth in the hands of the top 10% of households. Similarly, the difference  $\Delta_{HW} = s_{PI}^{10}(t) - \widetilde{s}_{PI}^{10}(t, \delta_{HW})$  quantifies the role of human wealth hoarding by the top households. Finally, the difference  $\Delta_{resid} = (s_{PI}^{10}(t) - s_{PI}^{10}(1989)) - \Delta_{NW} - \Delta_{HW}$  identifies how much of the change in PI concentration is due to a shift in the composition of PI towards  $NW$  or  $HW$ , rather a change in the marginal distributions of  $NW$  or  $HW$ .

We use this decomposition to make sense of changes in PI concentration between 1989

---

<sup>20</sup>It can be shown that

$$\delta_x(t) = x^{10}(t) \cdot \left[ \frac{x^{90}(1989)}{x^{10}(1989) + x^{90}(1989)} \right] - x^{90}(t) \cdot \left[ \frac{x^{10}(1989)}{x^{10}(1989) + x^{90}(1989)} \right].$$

and 2013. As shown in Table 2 the share of permanent income in the hands of the top 10% of households (ranked according to their net worth) went up from 0.283 to .395. That is, in 2013 the the top 10% of households by net worth managed to lay claims on an extra 11% of total resources in the economy. Out of this 11% gain, roughly 3.6% was due to higher concentration in the marginal distribution of net worth, while only 0.2% can be attributed to more concentration in the marginal distribution of human wealth. The remaining change (roughly 7.4% out of 11%) can be attributed to (i) a change in the composition of permanent income that puts more weight on net worth and less on human wealth, and/or (ii) to an increase in the share of households that sit at the top of both the net worth and human wealth distribution. However, the share of households who belong to the top 10% of both marginal distributions (of net worth and human wealth) actually decreased slightly from 16.6% in 1989 to 15.1% in 2013. This implies that the higher concentration of permanent income is mostly due to the increasing importance of real/financial net worth as a component of permanent income.

While different families populate the top of the distributions of net worth and human wealth, the role of asset wealth as a driver of permanent income appears to have increased significantly between 1989 and 2013, and this largely explains the higher concentration of permanent income in the hands of high net worth households.

## **6 Conclusions**

Accounting for heterogeneity in wealth and lifetime resources is key to provide a broader assessment of cross-sectional inequality and of its evolution. In this paper we outline a new approach that allows to quantify the value of human capital (human wealth) held by different households. Our analysis brings together different data sources and delivers estimates of households' wealth and permanent income. These estimates do not require strong assumptions about preferences or income processes as they rely on novel results on non-parametric

identification. Our estimates contain new and valuable information about heterogeneity along a variety of wealth and income measures. This information is especially useful when accounting for the changing patterns of wealth inequality over the past three decades.

We show that human wealth is less concentrated than net worth. Hence, inequality in permanent income is actually lower than inferred from popular measures of inequality that only focus on asset wealth. However, it is also apparent that richer households have accrued a growing share of permanent income. In fact, concentration of permanent income has grown much faster than concentration of net worth. As a consequence, effective inequality has grown more than previously thought, albeit from a lower initial level.

We document that changes in the marginal distributions of net worth and human wealth only account for a small part of the significant increase in permanent income concentration. Through simple accounting exercises we show that the increasing concentration of permanent income is mostly due to the mounting importance of asset wealth as a share of total wealth. We also find that the share of households who sit at the top of *both* net worth and human wealth distributions has actually decreased between 1989 and 2013, indicating that increased concentration of permanent income cannot be explained by a small set of households hoarding all types of wealth. Instead, the key driver of permanent income concentration seems to be the expansive growth of real/financial assets as share of the wealth portfolios of rich households. High net worth households, rather than high human wealth households, account for a larger share of total permanent income in 2016 than they did in 1989, suggesting that changes in wealth composition may be key to understand recent inequality patterns.

## References

- Athreya, K., Ionescu, F., and Neelakantan, U. (2015). ‘Stock market participation: The Role of Human Capital.’ Tech. rep., Working Paper, Federal Reserve Bank of Richmond.
- Athreya, K. B., Ionescu, F., and Neelakantan, U. (2017). ‘College or the Stock Market, or College and the Stock Market?’ Tech. rep., Board of Governors of the Federal Reserve System (US).
- Attanasio, O. and Pistaferri, L. (2014). ‘Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure.’ *The American Economic Review*, *PP*, vol. 104(5):pages 122–126.
- Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). ‘Trends in US Wage Inequality: Revising the Revisionists.’ *The Review of Economics and Statistics*, vol. 90(2):pages 300–323.
- Becker, G. S. (1975). *Human Capital*. Second Edition, National Bureau of Economic Research, New York, NY.
- Benzoni, L. and Chyruk, O. (2015). ‘The Value and Risk of Human Capital.’ Working paper, Chicago Fed.
- Bonhomme, S., Lamadon, T., and Manresa, E. (2017). ‘Discretizing unobserved heterogeneity.’ Tech. rep., University of Chicago.
- Bricker, J., Henriques, A., Krimmel, J., and Sabelhaus, J. (2016). ‘Measuring Income and Wealth at the Top using Administrative and Survey Data.’ *Brookings Papers on economic Activity*.
- Conesa, J. C., Kitao, S., and Krueger, D. (2009). ‘Taxing capital? Not a bad idea after all!’ *The American economic review*, vol. 99(1):pages 25–48.

- Cunha, F., Heckman, J., and Navarro, S. (2005). 'Separating Uncertainty from Heterogeneity in Life Cycle Earnings.' *Oxford Economic Papers*, vol. 57(2):pages 191–261.
- Davila, J., Hong, J. H., Krusell, P., and Ríos-Rull, J.-V. (2012). 'Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks.' *Econometrica*, vol. 80(6):pages 2431–2467.
- De Nardi, M., Fella, G., and Paz-Pardo, G. (2016). 'The Implications of Richer Earnings Dynamics for Consumption and Wealth.' NBER working paper no. 21917.
- DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). 'Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach.' *Econometrica*, vol. 64:pages 1001–1044.
- Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., and Srisuma, S. (2016). 'Nonparametric Euler Equation Identification and Estimation.' Tech. rep., mimeo, Boston College.
- Gallipoli, G. and Turner, L. (2011). 'Household Responses to Individual Shocks: Disability and Labour Supply.' Mimeo.
- Goldin, C. and Katz, L. F. (2007). 'The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005.' NBER Working Paper 12984.
- Golosov, M., Tsyvinski, A., Werning, I., Diamond, P., and Judd, K. L. (2006). 'New Dynamic Public Finance: A User's Guide [with Comments and Discussion].' *NBER macroeconomics annual*, vol. 21:pages 317–387.
- Gottschalk, P., Moffitt, R., Katz, L. F., and Dickens, W. T. (1994). 'The growth of earnings instability in the US labor market.' *Brookings Papers on Economic Activity*, vol. 1994(2):pages 217–272.

- Guner, N., Kaygusuz, R., and Ventura, G. (2014). ‘Income taxation of U.S. households: Facts and parametric estimates.’ *Review of Economic Dynamics*, vol. 17(4):pages 559 – 581. ISSN 1094-2025.
- Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2016). ‘What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics?’ Working Paper, University of Minnesota.
- Heathcote, J., Perri, F., and Violante, G. L. (2010). ‘Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006.’ *Review of Economic Dynamics*, vol. 13(1):pages 15–51.
- Huggett, M. and Kaplan, G. (2016). ‘How large is the Stock Component of Human Capital?’ *Review of Economic Dynamics*, vol. 22:page 2151.
- Jorgenson, D. and Fraumeni, B. (1989). ‘The Accumulation of Human and Non-Human Capital, 1948-1984.’ In R. Lipsey and H. Tice, eds., *The Measurement of Saving, Investment and Wealth*. NBER Studies in income and Wealth.
- Kaymak, B. and Poschke, M. (2016). ‘The evolution of wealth inequality over half a century: The role of taxes, transfers and technology.’ *Journal of Monetary Economics*, vol. 77:pages 1–25.
- Kocherlakota, N. R. (2010). *The new dynamic public finance*. Princeton University Press.
- Levy, F. and Murnane, R. J. (1992). ‘US Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations.’ *Journal of Economic Literature*, vol. 30(3):pages 1333–1381.
- Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

- Lucas Jr, R. E. (1978). 'Asset prices in an exchange economy.' *Econometrica: Journal of the Econometric Society*, pages 1429–1445.
- Piketty, T. and Saez, E. (2006). 'The evolution of top incomes: a historical and international perspectives.' In *American Economic Review Papers and Proceedings*, pages 200–205. Volume 96, number 2.
- R. Haveman, R. Bershader, A. and Schwabish, J. (2003). 'Human Capital in the United States from 1975 to 2000: Patterns of Growth and Utilization.' Upjohn Institute.
- Rios-Rull, J.-V., Kuhn, M. *et al.* (2016). '2013 Update on the US Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics.' *Quarterly Review*, pages 1–75. April.
- Saez, E. and Kopczuk, W. (2004). 'Top Wealth Shares in the United States: 1916-2000: Evidence from Estate Tax Returns.' NBER Working Paper 10399.
- Saez, E. and Zucman, G. (2014). 'Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data.' NBER working paper 20625.

## A Derivation of the Pricing Equation

We derive an individual's valuation of his/her own human capital by determining the shadow price of an asset that exactly replicates that individual's state-contingent labor market outcomes. To accomplish this we introduce a hypothetical asset that pays dividends per share equal to individual  $i$ 's yearly labor income, but also requires the individual to commit to their state-contingent labor supply plan.<sup>21</sup> Because of this commitment we replace  $w_{it}h_{it}$  from the problems described above with  $y_{it}$ , with the understanding that  $y_{it}$  is state-contingent earnings under the optimal labor supply plans of problems (1) and (4) above.<sup>22</sup>

**Human Capital Valuations of Married Individuals.** We begin by valuing individual  $i$ 's human capital when  $i$  is married. The number of shares of the hypothetical asset that  $i$ 's household owns at time  $t$  is  $e_{it}$ , and the price of this asset is  $\theta_{it}$ . We could also introduce an asset based on  $j$ 's human capital, but that is not necessary to value  $i$ 's human capital, hence we suppress that notation for now. When the hypothetical asset  $e_{it}$  is introduced, the budget constraint for a married household becomes:

$$\sum_{\kappa \in k} a_{(ij)t+1}^{\kappa} + c_{(ij)t} + \theta_{it}e_{it+1} \leq \theta_{it}e_{it} + (1 + e_{it})y_{it} + y_{jt} \quad (39)$$

$$+ \sum_{\kappa \in k} R_t^{\kappa} a_{(ij)t}^{\kappa} - T_t(\mathbf{a}_{(ij)t}, y_{it}, y_{jt}).$$

Furthermore, we include  $e_{it}$  as an additional state variable in the household planner's problem in equation (4), as well as in the definition of an individual's utility from marriage in (3). Given

---

<sup>21</sup>Of course, in reality no one would be willing to buy this asset from  $i$  because of the inherent commitment problem. Hence, the valuation we derive is truly a shadow price representing what human capital is worth to its owners. As discussed at length by Benzoni and Chyruk (2015), it is not normally possible to enforce contracts written against future labor services and ownership of human capital is not transferable (that is, human capital is a non-traded asset).

<sup>22</sup>As noted by Huggett and Kaplan (2016), this approach to valuing non-traded assets was first introduced by Lucas Jr (1978). Huggett and Kaplan (2016) also adopt this approach.

these adjustments we can rewrite the household planner's problem in a recursive manner as:

$$\begin{aligned}
V_{(ij)}^M(\mathbf{a}_{(ij)t}, e_{it}, X_{it}, X_{jt}, \eta_i, \eta_j, \Omega^t) = & \quad (40) \\
\max_{\mathbf{b}_{(ij)t}} \left\{ \lambda_{(ij)} u(c_{it}, \ell_{it}) + (1 - \lambda_{(ij)}) u(c_{jt}, \ell_{jt}) \right. \\
& + \lambda_{(ij)} \beta (1 - \tilde{\mu}_{(ij)t}) E_{\{\Omega^{t+1}, \mathbf{a}_{it+1}\}} [V_i^S(\mathbf{a}_{it+1}, e_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1}) | \mathbf{a}_{(ij)t+1}] \\
& + (1 - \lambda_{(ij)}) \beta (1 - \tilde{\mu}_{(ij)t}) E_{\{\Omega^{t+1}, \mathbf{a}_{jt+1}\}} [V_j^S(\mathbf{a}_{jt+1}, X_{jt+1}, \eta_j, \Omega^{t+1}) | \mathbf{a}_{(ij)t+1}] \\
& \left. + \beta \tilde{\mu}_{(ij)t} E_{\{\Omega^{t+1}\}} [V_{(ij)}^M(\mathbf{a}_{(ij)t+1}, e_{it+1}, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1})] \right\},
\end{aligned}$$

where the decision vector  $\mathbf{b}_{(ij)t}$  now includes  $e_{it+1}$ . After using the budget constraint in (39) to substitute  $c_{it}$  out of the problem in (40), we can easily derive the following first-order condition for the optimal choice of  $e_{it+1}$ :

$$\begin{aligned}
u_c(c_{it}, \ell_{it}) \vartheta_{it} = & \quad (41) \\
& \beta (1 - \tilde{\mu}_{(ij)t}) \frac{\partial}{\partial e_{it+1}} E_{\{\Omega^{t+1}, \mathbf{a}_{it+1}\}} [V_i^S(\mathbf{a}_{it+1}, e_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1}) | \mathbf{a}_{(ij)t+1}] \\
& + \frac{1}{\lambda_{(ij)}} \beta \tilde{\mu}_{(ij)t} \frac{\partial}{\partial e_{it+1}} E_{\{\Omega^{t+1}\}} [V_{(ij)}^M(\mathbf{a}_{(ij)t+1}, e_{it+1}, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1})].
\end{aligned}$$

To proceed we must calculate the derivatives of the married and single continuation values using envelope conditions. For the married continuation value this involves straightforward differentiation of equation (40) with respect to  $e_{it}$ , noting that the  $c_{it}$  has been replaced by the budget constraint. The result is,

$$\begin{aligned}
\frac{\partial}{\partial e_{it+1}} V_{(ij)}^M(\mathbf{a}_{(ij)t+1}, e_{it+1}, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1}) = & \quad (42) \\
\lambda_{(ij)} u_c(c_{it+1}^M, \ell_{it+1}^M) \vartheta (\theta_{it+1}^M + y_{it+1}^M),
\end{aligned}$$

where the superscript  $M$  indicates quantities that arise during marriage. To obtain the derivative of a single person's value function we must first be explicit about the problem they solve when single. Extending equation (1) to include the hypothetical asset  $e_{it+1}$  results in the following problem:

$$\begin{aligned}
V_i^S(\mathbf{a}_{it}, e_{it}, X_{it}, \eta_i, \Omega^t) = & \tag{43} \\
\max_{c_{it}, \ell_{it}, h_{it}, \mathbf{a}_{it+1}} & \left\{ u(c_{it}, \ell_{it}) + \beta(1 - \mu_{it}) E_{\{\Omega^{t+1}\}} [V_i^S(\mathbf{a}_{it+1}, e_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1})] \right. \\
& \left. + \beta \mu_{it} E_{\{\Omega^{t+1}, X_{jt+1}, \eta_j, \mathbf{a}_{jt+1}\}} [V_i^M(\mathbf{a}_{(ij)t+1}, e_{it+1}, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1})] \right\}.
\end{aligned}$$

The maximization in (43) is subject to the usual time allocation and borrowing constraints, as well the extended budget constraint,

$$\begin{aligned}
\sum_{\kappa \in k} a_{it+1}^\kappa + c_{it} + \theta_{it} e_{it+1} \leq & \theta_{it} e_{it} + (1 + e_{it}) y_{it} \tag{44} \\
& + \sum_{\kappa \in k} R_t^\kappa a_{it}^\kappa - T_t(\mathbf{a}_{it}, w_{it}, h_{it}).
\end{aligned}$$

The derivative of the value function in (43) can thus be derived by replacing  $c_{it}$  with the extended budget constraint, resulting in:

$$\frac{\partial}{\partial e_{it+1}} V_i^S(\mathbf{a}_{it+1}, e_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1}) = u_c(c_{it+1}^S, \ell_{it+1}^S) (\theta_{it+1}^S + y_{it+1}^S). \tag{45}$$

Finally, using equations (42) and (45), one can re-arrange the first order condition for optimal  $e_{it+1}$  chosen by a married household (equation 41) into an expression describing the valuation

of  $i$ 's human capital  $\theta_{it}^M$  (the purchase price per share of  $e_{it+1}$ ):

$$\begin{aligned} \theta_{it}^M = & \beta(1 - \tilde{\mu}_{(ij)t}) \frac{1}{\vartheta} E_{\{\Omega^{t+1}, \mathbf{a}_{it+1}\}} \left[ \frac{u_c(c_{it+1}^S, \ell_{it+1}^S)}{u_c(c_{it}, \ell_{it})} (y_{it+1}^S + \theta_{it+1}^S) \right] \\ & + \beta \tilde{\mu}_{(ij)t} E_{\{\Omega^{t+1}\}} \left[ \frac{u_c(c_{it+1}^M, \ell_{it+1}^M)}{u_c(c_{it}, \ell_{it})} (y_{it+1}^M + \theta_{it+1}^M) \right]. \end{aligned} \quad (46)$$

The result that stochastic discount factors are a component of the value of human capital in this model is related to general asset pricing formulations found in the literature following the seminal work of Lucas (1978). The probability of a change in marital status, and the surplus generated by marriage (through the economies of scale parameter  $\vartheta$ ) also factor into our valuation results.

**Human Capital Valuations for Single Individuals.** We derive the human capital valuation equations of an unmarried individual by considering their first-order condition for the optimal choice of  $e_{it+1}$  in problem (43):

$$\begin{aligned} u_c(c_{it}, \ell_{it}) \theta_{it} = & \quad (47) \\ & \beta(1 - \mu_{it}) \frac{\partial}{\partial e_{it+1}} E_{\{\Omega^{t+1}, \mathbf{a}_{it+1}\}} \left[ V_i^S(\mathbf{a}_{it+1}, e_{it+1}, X_{it+1}, \eta_i, \Omega^{t+1}) | \mathbf{a}_{(ij)t+1} \right] \\ & + \beta \mu_{it} \frac{\partial}{\partial e_{it+1}} E_{\{\Omega^{t+1}, X_{jt+1}, \eta_j, \mathbf{a}_{jt+1}\}} \left[ V_i^M(\mathbf{a}_{(ij)t+1}, e_{it+1}, X_{it+1}, X_{jt+1}, \eta_i, \eta_j, \Omega^{t+1}) \right]. \end{aligned}$$

As was the case when deriving valuations for married individuals, we need to substitute out the derivatives of continuation values. For the derivative of  $V_i^S(\cdot)$  this is straightforward, and in fact we have the expression in equation (45) already. However, the derivative of  $V_i^M(\cdot)$  proves more difficult because we cannot resort to a standard envelope condition. This is the case because  $V_i^M(\cdot)$  is *not* an indirect utility function, or in other words is *not* the solution to an individual optimization problem. Rather,  $V_i^M(\cdot)$  is a component of the objective function maximized by the household planner. To compute the necessary derivative here we must first

characterize the effect of pre-marital investments on the utility allocated to the spouse making those investments, which requires us to make assumptions about how the Pareto weight  $\lambda_{(ij)}$  is determined in the event that  $i$  gets married. Indeed, valuation of pre-marital human capital investments is inextricably linked to the household bargaining process upon marriage.

As anticipated above, we assume symmetric Nash Bargaining over the surplus generated by marriage. Under this assumption we can derive a relationship pinning down how the marital utility of person  $i$  changes if they make pre-marital investments. Symmetric Nash Bargaining implies that  $i$ 's utility in marriage must increase by at least as much as their outside option (utility from being single), plus half of any surplus generated by pre-marital investment.

Specifically we assume that a married household's Pareto weight solves

$$\max_{\{V_i^M, V_j^M\}} (V_i^M - V_i^S) (V_j^M - V_j^S), \quad (48)$$

where we have suppressed the state variables within the value functions for clarity. Let  $G(V_i^M, V_j^M) = 0$  be the Pareto frontier of household allocations, in which case the Nash Bargaining solution must satisfy

$$(V_i^M - V_i^S) = \frac{G_2}{G_1} (V_j^M - V_j^S). \quad (49)$$

To translate this condition into something empirically useful, note that an equivalent formulation of the household planning problem in equation (40) is:

$$\max \{ \lambda_{(ij)} V_i^M + (1 - \lambda_{(ij)}) V_j^M \}$$

subject to

$$G(V_i^M, V_j^M) = 0.$$

Combining the first-order conditions from this problem with those from the underlying Nash

Bargaining problem results in:

$$(V_i^M - V_i^S) = \frac{1 - \lambda_{(ij)}}{\lambda_{(ij)}} (V_j^M - V_j^S). \quad (50)$$

The equivalence of equations (49) and (50) is due to the fact that  $\lambda_{(ij)}$  is the Pareto weight that implicitly solves the Nash Bargaining problem in equation (48).

Next, we examine equation (49) evaluated at the point at which person  $i$  brings exactly zero units of  $e_{it}$  to the marriage, as this is the solution we observe in the data. Computing the total differential of this equation with respect to  $e_{it}$  results in

$$\frac{\partial V_i^M}{\partial e_{it}} - \frac{\partial V_i^S}{\partial e_{it}} = \left( \frac{G_2}{G_1} \right) \frac{\partial V_j^M}{\partial e_{it}} + \frac{1}{G_1} \left( \frac{\partial G_2}{\partial e_{it}} (V_j^M - V_j^S) - \frac{\partial G_1}{\partial e_{it}} (V_i^M - V_i^S) \right). \quad (51)$$

While this expression may seem intractable, one can easily show that at the optimal solution to the household planner's problem

$$\left( \frac{\partial G_2}{\partial e_{it}} / \frac{\partial G_1}{\partial e_{it}} \right) = \frac{u_c(c_{it}, \ell_{it})}{u_c(c_{jt}, \ell_{jt})} = \frac{\lambda_{(ij)}}{1 - \lambda_{(ij)}}. \quad (52)$$

Therefore, the last term of equation (51) equals zero when evaluated at the solution to the bargaining problem. Thus, a final simplified relationship between the derivatives of individual utilities, evaluated at the solution to the bargaining problem, is

$$\frac{\partial V_i^M}{\partial e_{it}} - \frac{\partial V_i^S}{\partial e_{it}} = \frac{1 - \lambda_{(ij)}}{\lambda_{(ij)}} \frac{\partial V_j^M}{\partial e_{it}}. \quad (53)$$

Intuitively, the extent to which  $i$ 's utility in marriage will increase in excess of their outside option depends on their ex-post Pareto weight and how valuable the hypothetical asset would be to their spouse.

To utilize equation (53), first note that the definition of the household planner's optimiza-

tion objective in (4) implies that the envelope condition in (42) can be re-written as:

$$\lambda_{(ij)} \frac{\partial V_{it+1}^M}{\partial e_{it+1}} + (1 - \lambda_{(ij)}) \frac{\partial V_{jt+1}^M}{\partial e_{it+1}} = \lambda_{(ij)} u_c(c_{it+1}^M, \ell_{it+1}^M) \vartheta (\theta_{it+1}^M + y_{it+1}^M). \quad (54)$$

Combining this with the Nash Bargaining implication in (53), we obtain an extremely useful result characterizing the effect of pre-marital investments on the utility within marriage:

$$\frac{\partial V_{it+1}^M(\cdot)}{\partial e_{it+1}} = \frac{1}{2} u_c(c_{it+1}^M, \ell_{it+1}^M) \frac{1}{\vartheta} (\theta_{it+1}^M + y_{it+1}^M) + \frac{1}{2} \frac{\partial V_{it+1}^S(\cdot)}{\partial e_{it+1}}. \quad (55)$$

The intuition for this equation relates to how much of the return on the hypothetical asset will be allocated to individual  $i$  by the household planner. A lower bound is the change in their utility if they exercise their outside option, which is captured by  $\partial V_{it+1}^S / \partial e_{it+1}$ . An upper bound is the marginal change in their utility if the entire return on the asset, including surplus due to economies of scale, is allocated to  $i$ . With symmetric bargaining exactly one half of the component pertaining to the return that exceeds the effect on  $i$ 's outside option is paid to  $i$ .

Equation (55) is useful because we now have an expression to substitute into equation (41), which was our objective when we set out to analyze the bargaining problem. Doing this, and substituting the envelope condition for single households in equation (45), allows us to derive the following valuation formula for the human capital of a currently unmarried person  $i$ :

$$\begin{aligned} \theta_{it}^S = & \beta \left(1 - \frac{\mu_{it}}{2}\right) E_{\{\Omega^{t+1}\}} \left[ \frac{u_c(c_{it+1}^S, \ell_{it+1}^S)}{u'(c_{it}, \ell_{it})} (y_{it+1}^S + \theta_{it+1}^S) \right] \\ & + \beta \frac{\mu_{it}}{2} E_{\{\Omega^{t+1}, X_{jt+1}, \eta_j, \mathbf{a}_{jt+1}\}} \left[ \frac{u_c(c_{it+1}^M, \ell_{it+1}^M)}{u'(c_{it}, \ell_{it})} \frac{1}{\vartheta} (y_{it+1}^M + \theta_{it+1}^M) \right]. \end{aligned} \quad (56)$$

While this expression is similar to canonical asset pricing formulations, it makes clear that the correct pricing relationship involves a biased expectation of future returns to human capital, where the bias derives from the implicit extra weight single households place on outcomes in the event of remaining single. The above equation is also informative as to how one would

test the robustness of the symmetric bargaining assumption: asymmetric bargaining weights would result in factors other than 1/2 (but still on the unit interval) being used to re-weight single and married outcomes.

We can subsume all sources of uncertainty into a single expectation operator  $\mathbb{E}_{it}$ , which also accounts for the re-weighting of unmarried future outcomes (as opposed to an unweighted expectation  $E_{it}$ ). Having done this we can summarize the value of human capital for any individual as

$$\theta_{it} = \mathbb{E}_{it} \left[ \beta \frac{u_c(c_{it+1}, \ell_{it+1})}{u_c(c_{it}, \ell_{it})} (y_{it+1} + \theta_{it+1}) \right], \quad (57)$$

where future variables implicitly depend on marital status. Clearly, valuations of one's own human capital depend on stochastic discount factors. Thus, state-contingent realizations of individual consumption matter for valuing state-contingent human capital payoffs. The last step in our analysis is to evaluate equation (57) at the point  $e_{it} = 0$  so that the equation is analogous to real-world valuations where human capital assets are not traded. Then, given some estimate of the distribution of state-contingent consumption realizations and appropriate weighting of future outcomes, human capital valuations can be estimated.

## B Supplementary Figures

Figure 11: Average human wealth, net worth and total wealth over the lifecycle, across their respective percentiles. Values in 2016 dollars.

