INTERTEMPORAL LABOR SUPPLY AND HUMAN CAPITAL ACCUMULATION *

BY SUSUMU IMAI AND MICHAEL P. KEANE†

Department of Economics, Concordia University, and CIREQ, Canada; Department of Economics, Yale University

We solve and estimate a dynamic model that allows agents to optimally choose their labor hours and consumption and that allows for both human capital accumulation and savings. Estimation results and simulation exercises indicate that the intertemporal elasticity of substitution is much higher than the conventional estimates and the downward bias comes from the omission of the human capital accumulation effect. The human capital accumulation effect renders the life-cycle path of the shadow wage relatively flat, even though wages increase with age. Hence, a rather flat life-cycle labor supply path can be reconciled with a high intertemporal elasticity of substitution.

1. INTRODUCTION

The intertemporal elasticity of substitution in labor supply (i.e.s.) has been a topic of considerable interest in both labor and macroeconomics for at least the past 30 years (see, e.g.,**?**). Recently, there have been several studies that address the question using micropanel data. Classic examples are **?**, **?**, and **?**. They focus on estimating the intertemporal elasticity of substitution in labor supply, using marginal utility of wealth constant labor supply functions. In their work they assume that the utility function is time separable and wages are exogenous.

But if current labor supply leads to human capital accumulation (i.e., learning by doing), then estimates of the i.e.s. under a false assumption of no human capital accumulation are likely to be biased towards zero. The reason is as follows: As the wage increases over the life cycle, the substitution effect induces labor supply to increase, thus providing an incentive for people to supply more labor in older age. On the other hand, both concavity of the value function with respect to human capital and the approaching retirement period lower the marginal rate of return to human capital investment, thus reducing the incentive to supply labor. If these two factors roughly cancel,

*Manuscript received May 2000; revised December 2

†We are grateful to John Geweke, Richard Rogerson, Chris Flinn, the editor, and an anonymous referee for many useful comments. Please address correspondence to: Susumu Imai, Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, QC, H3G 1M8, Canada. Phone: (514)-848-2424, extension 3907. Fax: (514)-848-4536. E-mail: simai@alcor.con

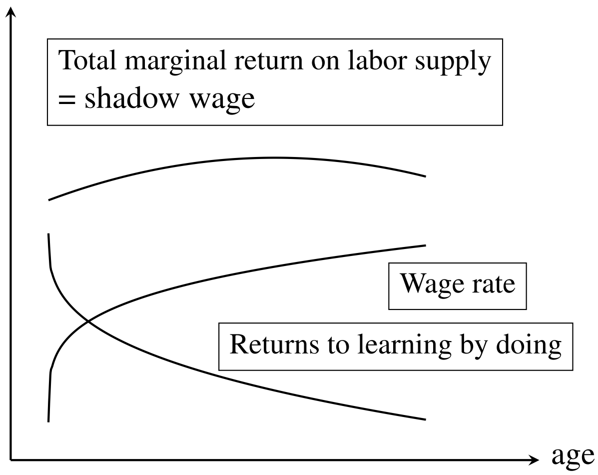


Figure 1: OPTIMAL LIFE CYCLE LABOR SUPPLY

then even if wages increase over the life cycle, labor supply will be little changed (see Figure 1). If we only allow for the substitution effect and not the human capital effect, the i.e.s. is identified primarily from the covariation of the wages and hours over the life cycle. Then, we will falsely conclude that the i.e.s. is low, simply because labor supply remains roughly constant over the life cycle even though wages increase[2].

In this article, we address the issue of human capital accumulation in two steps. First, we estimate the life cycle labor supply model using maximum likelihood (ML) estimation based on a full solution of agents' dynamic programming problem that allows human capital accumulation. In the estimation, we use the white male sample in the NLSY79 data (see Section 5). Our estimate of the disutility of labor or parameter implies that the i.e.s. is 3.82, which is quite comparable to results obtained and used in the macroliterature. What drives this result is that once human capital is included in the model, the i.e.s. is identified off the covariation of hours with the opportunity cost of time (not just the wage rate). Since the life cycle pattern of the opportunity cost of time is fairly flat, the i.e.s. is identified primarily off of short-run covariation between hours and wages.

In the second step, we simulate data from the estimated model. Using the simulated data, we estimate consumption and labor supply Euler equations like those in **?** and **?**, which do not allow for human capital accumulation. The elasticity estimates we obtain from the simulated data using the MaCurdy estimation method range from 0.3 to 1.1 when we use all the simulated data, and from 0.1 to 1.7 after we remove the outliers from the simulated data using a similar procedure to **?**. The elasticity estimates obtained using the consumption and labor supply data as in **?** range from -0.2 to 2.8. These estimates are significantly lower than the ML estimate. OLS and IV results from the NLSY97 data are also reported and the i.e.s. is again estimated to be low.

The high elasticity obtained by full solution estimation, and the contrasting low elasticities implied by the conventional estimation methods imply that the latter are significantly biased towards

[2]**?** makes a similar point in the context of a model where workers choose the fraction of time on the job to devote to investment in human capital (in contrast to our learning-by-doing setup). Since workers are only paid for time spent in production, "measured wage rates obtained by dividing weekly income by weekly reported hours on the job systematically understate the true wage rate." This understatement of the "true wage rate" (or opportunity cost of time) is greatest at younger ages. Thus, over the life cycle the opportunity cost of time is flatter than the wage rate

zero, and that one of the main reasons behind this is the omission of human capital accumulation. Thus, our results may explain the apparent contradiction between the macro- and microliterature noted above.

Notice that in conventional methods of estimation such as **?** or **?**, the i.e.s. is defined and estimated as the elasticity of substitution when workers change labor supply along the anticipated life cycle wage path. But in macroeconomics, the discussion is typically about how labor supply responds to unanticipated business cycle shocks. In our analysis, we explicitly solve the dynamic programming model including unanticipated wage shocks. Therefore, our estimate is more relevant in providing microevidence for use in calibrating real business cycle models. After the estimation, we also simulate the hours response to a temporary 2% increase in the rental rate on human capital. There, we show that even though the elasticity of labor supply is 3.82, this does not imply that hours increase by roughly 4 times the percentage amount of the wage increase. Rather, although such a large response is observed for older workers, the response for young workers is much smaller. This is because of the role of human capital accumulation, which we discuss more later [3].

The theory of optimal life-cycle consumption and labor supply with human capital accumulation has been developed by **?** and **?**, **?**, **?**, and others. They provide a neoclassical theoretical framework to explain the life-cycle profiles of wages, schooling, and work hours. **?** was among the first to estimate a dynamic labor supply model that includes human capital accumulation. **?** estimated Euler equations of optimal consumption and labor supply using the nonlinear GMM method. But in her work, she used the translog utility function for consumption and labor supply hours. Hence, she did not explicitly estimate the i.e.s. parameter. **?** also estimate a dynamic labor supply model for females.

? also include human capital in a life cycle labor supply model that they estimate by ML. But they restrict labor hours to two categories: zero hours or full time. If people intertemporally substitute labor but still stay within the same broad hours classification, their estimate of the i.e.s. may be downwardly biased. But if most people work around the borders of the two categories, then small changes in labor supply will be classified as changes between two categories; thus the i.e.s. will be overestimated. Since the former case seems to be more likely in a discrete choice model of labor supply, we suspect that one is likely to get downwardly biased estimates of the i.e.s.. Also, since they use a linear utility function for consumption, there is no wealth effect in their model, even though the wealth effect may be an important factor linking peoples' decisions intertemporally. **?** estimate a model with human capital accumulation and saving, but hours are also discrete in their model (e.g., full time, part time, and zero hours work).

The major obstacle to ML estimation of the dynamic labor supply model with continuous hours and human capital accumulation is that the full solution of the continuous variable dynamic programming problem implied by this model is extremely computationally demanding. There are two reasons. First, the state space of the dynamic programming problem is now infinite. Even in a discrete choice dynamic programming problem, where there is only a finite state space, researchers

[3] Recently, several authors, such as **?** or **?**, have introduced learning by doing (LBD) into the standard real business cycle model as an internal propagation mechanism. **?** assume LBD in organizational capital at the plant level and **?** assume LBD in the human capital accumulation of workers over the life cycle. Their models succeed in generating persistence in macroeconomic variables even though their underlying technology shocks are set to be uncorrelated over time. Furthermore, their calibrated impulse-response function of output has the familiar hump-shaped form seen in the U.S. GDP time series data. Since the agents in our model are very similar to those assumed by **?**, our article provides some empirical evidence supporting the above line of research.

are usually plagued with the problem of having too many state space points to evaluate the value function (see, e.g., **?**; **?**). In the continuous choice case, explicit evaluation of the value function at each state space point is impossible. Second, in contrast to a discrete choice dynamic programming problem where solving for the control variables is a trivial optimization over a finite set of choices, in the continuous choice problem, solving for the control variables is the main source of computational burden. It requires a two-dimensional nonlinear Newton search algorithm to find optimal consumption and labor supply at each state space point.

Here we develop an algorithm that approximates the solution to the DP problem. The algorithm successively solves the Bellman equation backwards from the last period. First, we choose a finite set of grid points over assets and human capital at which to evaluate the expected value function (emax function). The emax function derived above is used for the next Bellman backward iteration. The main feature of the algorithm is that we avoid two-dimensional quadrature integration over both taste shocks and wage shocks by exploiting the fact that there is a one-to-one mapping from human capital to wages. Reducing the dimensionality of quadrature integration decreases the number of computationally demanding Newton searches by an order of magnitude. Also, by reducing the range of human capital points at which the value functions must be evaluated, it makes the Newton search algorithm itself easier and more accurate.

We apply our model to white males from the 1979 youth cohort of the National Longitudinal Survey of Labor Market Experience (NLSY79). Among the features of the NLSY79, which distinguishes it from the commonly used Panel Study of Income Dynamics (PSID), is detailed asset data for individuals from 1985. Instead of using food consumption to derive the marginal value of wealth at some period, which is commonly done by researchers using the PSID, we use asset data directly. We then derive total consumption by using the asset data and the intertemporal budget constraint.

The organization of the article is as follows. Section 2 presents the model, Section 3 describes the algorithm for solving the DP problem, and Section 4 describes the algorithm for forming the likelihood function. Section 5 describes the data, and Section 6 discusses the estimates and some model simulations. Section 7 concludes.

2. THE MODEL SPECIFICATION

In this section, we present a life-cycle model for an individual agent who rationally chooses his optimal life-cycle path for consumption and hours of labor supply.

At a given calendar time period s and age t, the agent's period utility of consumption is a concave function of the consumption of market goods, C, and the disutility of labor is a convex function of the hours of labor supply h and the taste shock ε_2. Preferences are additively separable over time. Agents choose optimal consumption and labor supply by maximizing their discounted expected life-cycle utility over the working horizon T, which is

$$E_t \sum_{\tau=t}^{T} \beta^{\tau}[u(C_\tau,\tau) - v(h_\tau \varepsilon_{2,\tau})] \tag{1}$$

Agents also face an intertemporal budget constraint and a human capital production constraint. The intertemporal budget constraint is

$$A_{t+1} = (1+r)A_t + W_{t,s}h_t - C_t \tag{2}$$

where A_t is the agent's asset holdings at age t and r is the interest rate. The observed wage $W_{t,s}$ at age t, time s is defined as the product of the human capital stock K_t times the rental rate on a unit of human capital, R_s

$$W_{t,s} = R_s K_t \tag{3}$$

The rental rate R_s is the market price of services of a unit of human capital. We assume a perfect market for human capital. Hence, at any time s, all agents face the same rental rate R_s.

Human capital evolves according to the human capital production function, which is a deterministic function of current labor supply hours h_t, current human capital K_t, and age t, along with the multiplicative wage shock, ε_t. That is,

$$K_{t+1} = g(h_t, K_t, t)\varepsilon_{1,t+1} \tag{4}$$

K_{t+1} is the age $t+1$ human capital after the wage shock $\varepsilon_{1,t+1}$ is realized.

At age t, time period s, the agent's decision process can be described by the following maximiation of the value function.

$$\begin{aligned} V_{t,s}(A_t, K_t, \varepsilon_{2,t}) = \max\nolimits_{C_t,h_t} \{ & u(C_t,t) - v(h_t, \varepsilon_{2,t}) \\ & + \beta E_t V_{t+1,s+1}[(1+r)A_t + R_s K_t h_t \\ & - C_t, g(h_t, K_t, t)\varepsilon_{1,t+1}, \varepsilon_{2,t+1}]\} \end{aligned} \tag{5}$$

For the utility function for consumption C, we choose a CRRA form augmented to include age effects

$$u(C_t, t) = A(t)\frac{C_t^{a_1}}{a_1}$$

where $A(t)$ is a spline in age, and $a_1 < 1$ is a constant.[4]

The disutility of labor, which is a function of hours h, is assumed to have the following functional form:

$$v(h_t, \varepsilon_{2,t} = \varepsilon_{2,t} b \frac{h_t^{a_2}}{a_2} \tag{6}$$

where $b > 0$ and $a_2 > 1$ are constants. Except for the added age effects in the marginal utility of consumption, the functional forms are adopted from **?** and **?**.[5] That will enable us to compare

[4]The reason that age effects are needed to explain observed consumption behavior is as follows: In the data, wages and labor supply are both relatively small on average when people are young compared to when they are at prime age. Hence, annual labor income is low when people are young. Therefore, if people smooth consumption, they should be in debt when young, and repay it in later years of the life cycle. But this is not the case in the data. (See Table 3 for mean age profiles of wage, hours, and assets.) Thus, age effects in consumption are necessary to explain the positive asset holdings for youths observed in the data. To capture this, the utility of consumption should be smaller when young, so that consumption rises over time, and individuals would not go heavily into debt early in life. Alternative mechanisms to explain the observed asset pattern would be liquidity constraints, as in **?**, or a more general utility function with a very strong precautionary saving motive.

The age effect $A(t)$ starts at $C_0 C_1$ at age 20, then gradually changes to $C_0 C_2$ at age 25 and to C_0 at age 33. Thereafter, $A(t)$ stays constant at the value C_0. That is, $A(t)$ is a linear spline with kinks at age 25 and 33. Agents in our model attempt to equate the marginal utility of consumption across time. If they place less value on consumption when young, this would be reflected in $C_1 < 1$ and $C1 + C2 < 1$. Then, at younger ages, less consumption is necessary to reduce the marginal utility of consumption to a given level.

[5]By introducing intratemporal nonseparability of consumption and labor supply, one could explain the consumption profiles in the data without resorting to the age effects in consumption. But that would make the results of the estimation and simulation exercises less comparable to the results by **?** and **?**.

our results and theirs in later estimation and simulation exercises. Furthermore, in this parameter specification, the degree of intertemporal substitution of labor supply can be summarized by a single parameter, which is

$$i.e.s. \equiv b_2 \equiv \frac{1}{a_2 - 1}$$

Many empirical articles analyzing intertemporal labor supply behavior, such as **?** or **?**, use a translog function of consumption and leisure as the utility function. Although this approach has the advantage of being locally flexible, none of the parameters can be straightforwardly interpreted as describing the intertemporal elasticity of substitution in labor supply. Hence, from their estimation results, it is difficult to draw any conclusions about how much people substitute labor intertemporally, unless one simulates their estimated models.[6]

We assume the human capital production function in Equation (4) to be as follows:

$$g(K,h,t) = k_0 + \delta K + G(K,h,t) \tag{4a}$$

where $G(\cdot)$ is a function of current human capital, labor hours, and age t. Figure 2 gives some evidence on the shape of the human capital production function from the NLSY79 data. The figure shows the relationship between current labor supply hours and the next period hourly wage rate, within different cells for the current wage. Each cell has length 2 dollars. In light of this evidence, the $G(\cdot)$ component of the human capital production function is specified as follows:

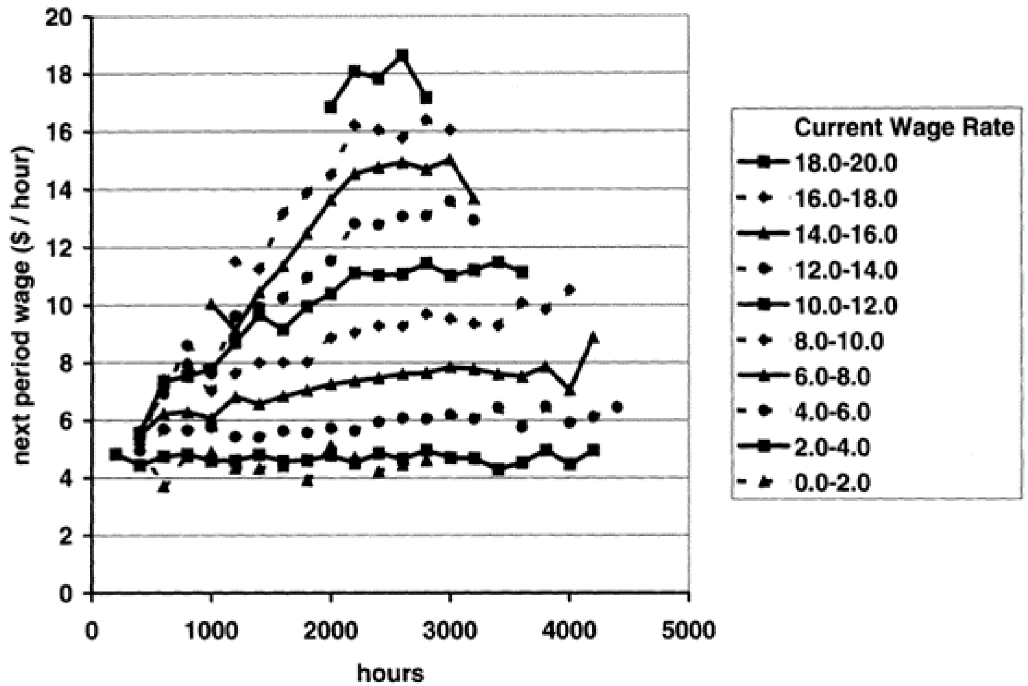
$$G(K,h,t) = A_0(1 + A_1(t-19))(B_1 + K)[(h+d_1)^{\alpha} - B_2(h+d_1)] \tag{4b}$$

This is designed to take account of the following features of the relationship between current hours, current wage rate, and the next period wage rate:

1. . The relation of future wages to current labor hours has a higher slope when the current wage is higher. That implies there is a significant complementarity between current wages and current hours in terms of learning by doing, which is captured by the multiplicative term $(B_1 + K)$.

2. The derivative of the human capital production function with respect to hours around $h = 0$ appears to be bounded. We capture that by introducing the intercept term d_1.

[6] Furthermore, in contrast to our full solution method, they use an Euler equation GMM method of estimation. There are some shortcomings of GMM estimation that are particularly related to the estimation of an intertemporal labor supply model that explicitly incorporates human capital accumulation, such as **?**. That is, potential nonconcavities in agents' problem as a result of human capital investment may create problems for GMM. If we think of labor supply as an input to income production, then an increase in labor supply by A percent in every period results in an increase in wage income of more than A percent. That is because an increase in labor supply also raises future wages via human capital accumulation. Hence, there are increasing returns to scale in the income-generating process and potential nonconvexities in the model. In such a case, just looking at the first-order conditions may not be sufficient to claim that agents are solving the intertemporal optimal labor supply problem. In our estimation process, we explicitly solve for the continuous variable dynamic programming problem and embed the solution in the ML estimation. Hence, our solution truly assumes that individuals choose optimal labor supply and consumption over the life cycle. Furthermore, the full solution method leads to straightforward simulation of the model after the estimation. The main goal of this research is to simulate the intertemporal labor supply model to conduct some estimation exercises to highlight the potential bias of the conventional estimation methods.



NOTE: Only plot cells with at least 20 observations.

Figure 2: HUMAN CAPITAL PRODUCTION FUNCTION

3. For very large hours, the slopes of the relation between future wages and current hours seems to be close to zero or even negative. Thus, there is a possibility that for very large hours, increase in hours have no effect on future human capital or even decrease future human capital. We can account for this possibility by adding the term $-B_2(h+d_1)$.

For estimation, we also include a pure age effect in the human capital production function, similar to the age effect included in the human capital earnings function by **?**, which corresponds to the term $A_1(t-19)$. The wage and taste shocks are assumed to have i.i.d. mean one log normal distributions. That is,

$$\ln(\varepsilon_i) \sim N(-\frac{1}{2}\sigma_i^2, \sigma_i), \quad i = 1,2 \tag{7}$$

We also allow for the measurement errors in wages, labor supply hours, and assets. We defer the discussion of the measurement error functional forms until Section 4.

We set the working horizon T at age 65. At the terminal period, we assume that agents get positive values from holding assets. This would arise, for example, because they are able to enjoy consumption until their death, and possibly can leave bequests to their heirs. We choose the following parameterization for the terminal value function:

$$V_{T+1}(A_{T+1}) = \begin{cases} 3\log(A_{T+1}+\phi) - 1 - 3\log(\phi) & \text{if } A_{T+1} > 0 \\ (\frac{A_{T+1}-\phi}{\phi})^3 & \text{otherwise} \end{cases} \tag{8}$$

where ϕ is a parameter that determines the marginal value of assets (at $T+1$) at various asset levels. Higher values of 0 imply that agents care less about the terminal assets. We chose this specification because this function is continuously differentiable in assets and the derivative is decreasing in assets. [7] It turns out that the coefficient ϕ is difficult to estimate (i.e., the likelihood is very flat over a wide range of 0), because the NLSY79 only has data on individuals until the age 36. After some experimentation, we set 0 to be 100,000. [8]

Now, to understand the effect of introducing human capital accumulation on the hours response to wage changes, we consider the first-order conditions of the above problem with respect to consumption and labor. These are:

$$\begin{aligned} C_t :& u_C(C_t,t) - \beta E_t V_{A,t+1,s+1}(A_{t+1},K_{t+1},\varepsilon_{2,t+1}) = 0 \\ h_t :& -v_h(h_t,\varepsilon_{2,t}) + R_s K_t u_C(C_t,t) \\ & + \beta E_t g_h \varepsilon_{1,t+1} V_{K,t+1,s+1}(A_{t+1},K_{t+1},\varepsilon_{2,t+1}) = 0 \end{aligned} \tag{9}$$

Notice that the current marginal disutility of labor equals the wage (R_sK_t) times the marginal utility of consumption, which is the marginal return to increases in current wage income due to increases in labor supply, plus an extra term that captures the marginal return to increases in future human capital. As the wage increases over the life cycle, the substitution effect induces labor supply to increase, thus providing an incentive for people to supply more labor in older age. This corresponds to the term $R_sK_tu_C(C_t,t)$. On the other hand, both concavity of the value function with respect to human capital and the approaching retirement period lower the marginal rate of return to human capital investment, thus reducing the incentive to supply labor. This comes from the term $\beta E_t g_h \varepsilon_{1,t+1} V_{K,t+1,s+1}(A_{t+1},K_{t+1},\varepsilon_{2,t+1})$. If these two factors roughly cancel, then even if wages increase over the life cycle, labor supply will be little changed (see Figure 1).

Observed heterogeneity is introduced into the model by allowing the parameters b, C_0, C_1, C_2 in preferences, and K_0, δ, A_0, A, and α in the human capital production function, to differ depend-

[7] The function is continuous at $A_{t+1} = 0$ since

$$\lim_{A_{T+1}\downarrow 0} V_{T+1}(A_{T+1}) = 3\log(0+\phi) - 1 - 3\log\phi = -1$$

and

$$\lim_{A_{T+1}\uparrow 0} V_{T+1}(A_{T+1}) = (\frac{0-\phi}{\phi})^3 = -1$$

Furthermore, the function is continuously differentiable at $A_{T+1} = 0$ since for $A_{T+1} > 0$,

$$\frac{\partial V_{T+1}(A_{T+1})}{\partial A_{T+1}}|_{A_{T+1}>0} = \frac{3}{A_{T+1}+\phi} \to \frac{3}{\phi}$$

as $A_{T+1} \searrow 0$ and

$$\frac{\partial V_{T+1}(A_{T+1})}{\partial A_{T+1}}|_{A_{T+1}\le 0} = 3(\frac{1}{\phi})^3(A_{T+1}-\phi)^2 \to \frac{3}{\phi}$$

as $A_{T+1} \nearrow 0$. Since both

$$\frac{\partial V_{T+1}(A_{T+1})}{\partial A_{T+1}}|_{A_{T+1}>0} = \frac{3}{A_{T+1}+\phi}$$

and

$$\frac{\partial V_{T+1}(A_{T+1})}{\partial A_{T+1}}|_{A_{T+1}\le 0} = 3(\frac{1}{\phi})^3(A_{T+1}-\phi)^2|_{A_{T+1}\le 0}$$

are decreasing in A_{T+1}, the derivative is decreasing in assets.

[8] Since we only have data until the age 36, and the final period is at age 65, our experimentation indicated that the likelihood function is very flat for a wide range of ϕ values around 100,000. But if ϕ is larger than 200,000, the lack of concavity of the final period value function causes the continuous variable DP solution algorithm to break down.

ing on whether an agent is a high school dropout, high school graduate, has some college, or is a college graduate. Completed schooling is treated as exogenous. We do not allow for unobserved heterogeneity, but, as we discuss in Section 6.4, the model is nevertheless able to generate substantial persistence in wages, hours, and assets due to the persistent nature of shocks to the human capital production function.

At this point, readers who are not interested in the algorithms for solving the dynamic programming problem and forming the likelihood can skip Sections 3 and 4 and go directly to the data description in Section 5.

3. SOLVING THE CONTINUOUS STOCHASTIC DYNAMIC PROGRAMMING PROBLEM

As discussed before, the problem agents solve in each period is as follows:

$$\begin{aligned} V_{t,s}(A_t, K_t, \varepsilon_{2,t}) = \max\nolimits_{C_t,h_t} \{ & u(C_t,t) - v(h_t, \varepsilon_{2,t}) \\ & + \beta E_t V_{t+1,s+1}[(1+r)A_t + R_s K_t h_t \\ & - C_t, g(h_t, K_t, t)\varepsilon_{1,t+1}, \varepsilon_{2,t+1}]\} \end{aligned} \tag{5'}$$

subject to the intertemporal budget constraint

$$A_{T_1} = (1+r)A_t + W_{t,s}h_t - C_t \tag{2}$$

and human capital production function

$$K_{t+1} = g(h_t, K_t, t)\varepsilon_{1,t+1} \tag{4}$$

where $\varepsilon_{1,t+1}$ is the human capital shock realized at age $t+1$.

Notice that the next period's human capital K_{t+1} is not known to the individual at age t. Let us rewrite the Bellman equation in terms of variables that the agent knows at age t. Define $\tilde{K}_{t+1}$ to be the next period human capital before the human capital shock is realized. That is,

$$\tilde{K}_{t+1} = g(h_t, K_t, t)$$

$$K_{t+1} = \tilde{K}_{t+1}\varepsilon_{1,t+1}$$

Define the value function in terms of $\tilde{K}_t$ as follows:

$$\begin{aligned} V_{t,s}(A_t, \tilde{K}_t, \varepsilon_{1,t}, \varepsilon_{2,t}) = \max\nolimits_{C_t,h_t} (& u(C_t,t) - v(h_t, \varepsilon_{2,t}) \\ & + \beta E_t V_{t+1,s+1}(A_{t+1}, \tilde{K}_{t+1}, \varepsilon_{1,t+1}, \varepsilon_{2,t+1})) \end{aligned}$$

Also, define the emax function V^E as follows"

$$V^E_{t+1,s+1}(A_{t+1}, \tilde{K}_{t+1}) = E_t V_{t+1,s+1}(A_{t+1}, \tilde{K}_{t+1}, \varepsilon_{1,t+1}, \varepsilon_{2,t+1})$$

Then, the above problem can be rewritten as follows:

$$V_{t,s}(A_t, \tilde{K}_t, \varepsilon_{1,t}, \varepsilon_{2,t}) = \max\nolimits_{C_t,h_t}[u(C_t,t) - v(h_t, \varepsilon_{2,t}) + \beta V^E_{t+1,s+1}(A_{t+1}, \tilde{K}_{t+1})]$$

subject to the intertemporal budget constraint (2), and human capital production function $\tilde{K}_{t+1} = g(h_t, K_t, t)$ where $K_t = \tilde{K}_t \varepsilon_{1,t}$.

There are several computational obstacles to solving the continuous stochastic dynamic programming problem that we assume the agents are facing. In order to numerically solve the above problem, in general, we have to start at the terminal period, T, and backsolve to $t = t_0$, where t_0 is the start of the planning period (assumed to be age 20).

Now, let the state space for the emax function be $(A_t, \tilde{K}_t)$. Suppose we have already solved for the emax function for age $t+1$. That is, we have already calculated the emax function

$$V^E_{t+1,s+1}(A_{t+1}, \tilde{K}_{t+1}) = E_t V_{t+1,s+1}(A_{t+1}, \tilde{K}_{t+1}, \varepsilon_{1,t+1}, \varepsilon_{2,t+1})$$

for all possible values of A_{t+1} and K_{t+1}. The next step in the backsolving process is to find the $V^E_{t,s}(A_t, \tilde{K}_t)$. Given the state space point, $(A_t, \tilde{K}_t)$, we need to derive the integral of $V_{t,s}(A_t, \tilde{K}_t, \varepsilon_{1,t}, \varepsilon_{2,t})$ with respect to $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$. Furthermore, in integrating for each value of the shock vector $(\varepsilon_{1,t}, \varepsilon_{2,t})$, we need to find the optimal consumption and labor supply to derive the value $V_{t,s}(A_t, \tilde{K}_t, \varepsilon_{1,t}, \varepsilon_{2,t})$.

To get an idea of the magnitude of the computational problem involved in solving this model, assume that agents only have two possible choices of consumption and labor supply, that is, a total of four discrete choices in each period. Also, suppose that there is no taste shock or human capital shock that we need to integrate over. Then, because the state variables in future periods depend on past choices, for a discrete choice dynamic programming problem with $T - 20$ time periods, we need to evaluate the value function at least 4^{T-20} state space points. For example, if we assume there are 20 time periods, this amounts to at least 1.099511D+12 points. Suppose on the other hand, we discretize the state space of assets and human capital into $n_A \times n_K$ grid points. Further suppose that at each grid point, we evaluate the value function with respect to $n_1 \times n_2$ combinations of human capital shocks and taste shocks, and integrate over the shocks to get the emax function. Then the required number of evaluations of the value function is at least $n_A x n_K x n_1 x n_2 x (T-20) x 4$, which is again extremely computationally demanding even with modest numbers of grid and quadrature points. In continuous choice dynamic programming problems, the state space is continuous, and hence the number of state space points is infinite. Therefore, explicit evaluation of the value function at each state space point (A_t, K_t) is impossible.

Furthermore, compared to discrete choice dynamic programming models, where optimization over the control variable only involves maximizing over a discrete set of choices, in the continuous choice problem we examine here, finding optimal consumption and labor supply requires a two-dimensional nonlinear search at each state space point.

In order to cope with the computational problem discussed above, we will use a set of approximation and interpolation methods. First, we only explicitly solve for the expected value functions at a finite set of asset and human capital grid points. The expected value functions at the remaining points are derived by Chebychev polynomial least squares interpolation. To solve for the expected value function at each state variable grid point, we need to integrate the value function over both the taste shock and the human capital shock. To avoid the quadrature integration with respect to the human capital shock, we add another interpolation and approximation step that exploits the fact that there is a one-to-one mapping from human capital to wages. We explain the algorithm in the appendix.

4. MAXIMUM LIKELIHOOD ESTIMATION

We use NLSY79 data to estimate the parameters of the model. There are several features of the data that we need to consider when we estimate the model. First, as in most other panel data, variables such as wages, labor supply, and assets are measured with error. Hence, the estimation procedure should incorporate a measurement error component in those variables. Second, there are periods where assets are missing. Hence, during the estimation process, we need to account for the missing asset data. In our likelihood function, we take into account both of these problems.

Suppose that for age t, period s, the true wage $W_{t,s} = R_s K_t$, labor supply, and assets $(W_{t,s}, h_t, A_t)$ are all observed with measurement error. Denote by $\xi_t = (\xi_{1,t}, \xi_{2,t}, \xi_{3,t})$ the vector of the measurement errors in observed labor income, hours of labor supply, and assets, respectively. Assume that the labor income measurement error is log normally distributed with mean 1. That is,

$$h_t^D = h_t + \xi_{2,t}$$
$$\ln(\xi_{1,t}) \sim N(-\frac{1}{2}\sigma_{\xi,1}^2, \sigma_{\xi,1})$$

where Y_t is the true labor income at period t (which equals $W_t h_t$) and Y_t^d is the observed labor income in the data.

Furthermore, for the measurement error in assets, we assume the following:

$$A_t^D = Ah_t + \xi_{3,t}$$
$$\xi_{3,t} \sim N(0, \sigma_{\xi,3})$$
$$\sigma_{\xi,3} = \sigma_{\xi,3,1} + \sigma_{\xi,3,2}(t-19)$$

where A_t is the true labor income at period t and h_t^d is the observed labor income in the data.

And in order to fill in the missing first period assets, we assume they are distributed as follows:

$$A_{t_0} \sim N(\bar{A}, V_{\bar{A}})$$

We let $\bar{A}$ differ depending on whether the first period is at age 20 or later. Also, we assume that

$$R_s = 1$$

for all periods s. Hence,

$$W_{t,s} = R_s K_t = W_t = K_t$$

We denote $K_t^D = W_t^D$, to be the observed human capital, which we derive by dividing the annual labor income by the annual hours worked, i.e.,

$$K_t^D = \frac{Y_t^D}{h_t^D}$$

This is different from the true human capital K_t by a measurement error component. That is, for observations after the initial period t_0,

$$K_t^D = K_t h_t \frac{\xi_{1,t}}{h_t^D}$$

Finally, for the initial period wage, we assume the following measurement error:

$$K_{t_0}^D = K_{t_0}\xi_0$$

where

$$\ln(\xi_0) \sim N(-\frac{1}{2}\sigma_{\xi_0}^2, \sigma_{\xi_0}^2)$$

Also, the interest rate r is set to be 5%.

Here, we adopt the simulated ML method. Denote by $\{K_t^m, h_t^m, C_t^m, A_t^m\}_{t=t_0}^T$ the sequence of the true human capital, true labor supply, true consumption, and true assets at the m^{th} simulation draw. We repeat the simulation M times and derive the likelihood in the steps described in the appendix.

As discussed earlier, the major obstacle to the ML estimation of the continuous choice dynamic programming problem is the iterative solution of the Bellman equation, which requires a Newton search routine for optimal consumption and labor supply at each asset and human capital grid point and quadrature point of the taste shock. In a standard ML routine, a single iteration requires evaluation of the likelihood and its partial derivatives with respect to all the model parameters.

The usual practice is to calculate the derivatives of the likelihood function numerically as follows: Suppose that the parameter vector is $\theta = (\theta_1, \theta_2, ..., \theta_n)$. Then, one solves the entire DP problem to evaluate the log likelihood for θ, $l(\theta, X^D)$, which is a function of parameters θ and data X^D. Then, for each $i = 1, ..., n$, one solves the DP problem and evaluates the likelihood at parameter values $(\theta_1, \theta_2, \ldots, \theta_i + \Delta_i, \ldots, \theta_n)$ where Δ_i is a small positive number. Then, the numerical derivative is

$$\frac{\partial l(\theta, X^D)}{\partial \theta_i} = \frac{l(\theta_1, \theta_2, \ldots, \theta_i + \Delta_i, \ldots, \theta_n) - l(\theta, X^D)}{\Delta_i}$$

Now, the effect on the value function of a marginal change in outside conditions can be decomposed into two components. The first component results from the change in the value function with the choice variables held constant, and the second component results from the change in the value function due to changes in the choice variables. From the envelope theorem, we know that the magnitude of the second component is of second order. Hence, as long as changes in the parameters are small, a valid approximation for the likelihood function under the parameter value

$$\theta'^i = (\theta_1, \theta_2, \ldots, \theta_i + \Delta_i, \ldots, \theta_n)$$

is obtained by constructing the value functions and the likelihood with the consumption and labor supply choices held fixed at the values derived under the parameter value θ. Because the approximation error in the value functions is of second order, the approximation error in the likelihood function evaluation is also of second order. Hence, for one evaluation of likelihood and all its partial derivatives, the Newton search algorithm over optimal consumption and labor supply at each grid point and quadrature point only needs to be done once. Since the Newton search algorithm is the most computationally demanding part of the whole likelihood evaluation, this significantly reduces the computational burden. In fact, this approach makes the computational cost of estimating the continuous choice model comparable to that of estimating discrete choice dynamic programming models.

5. DATA

The data are from the 1979 youth cohort of the National Longitudinal Survey of Labor Market Experience (NLSY79). The NLSY79 consists of 12,686 individuals, approximately half of them men, who were 14-21 years old as of January 1, 1979. The sample consists of a core random sample and an oversample of blacks, Hispanics, poor whites, and the military. One unique characteristic of the NLSY79 is that from 1985, it has comprehensive asset information for each respondent. In any intertemporal labor supply model, the shadow price of assets, or marginal utility of wealth, plays an important role as linking period-by-period decisions intertemporally. In the past, the Panel Study of Income Dynamics (PSID) was frequently used to estimate such models, and researchers either first differenced away the shadow price of assets, as in **?**, or used the marginal utility of food consumption as a proxy for the shadow price of assets, as in **?** or **?**. It was necessary for researchers analyzing the PSID to use food consumption data because that is the only consumption data it contains. Here we use the asset data directly to either measure the shadow price of assets, or, using the intertemporal budget constraint, back out total consumption.

We use the white male sample of the NLSY79 data. We only use males who are at least 20 years old and have completed schooling. In our analysis, we treat schooling as exogenous. Since people can either accumulate human capital by on the job experience or schooling, omission of the schooling decision can be an important source of bias. By only using data beginning from the year after the respondent last attended school, we hope to minimize the potential bias.[9] Also, we censor anybody who served in the military from the sample. Appendix A.3 describes in more detail how we constructed the data.

Since the NLSY79 only has asset data beginning in 1985, and the asset data in 1991 is missing, we recover the missing assets using the intertemporal budget constraint as discussed in the previous section. Table 1 gives the sample means of wages, hours of labor supply, and total wealth of individuals. Also, Table 2 gives the quantiles of the wage and labor supply distribution. Notice that the sample mean of the wages far exceeds the median. This indicates that there are some very high wage values. In order to remove the effect of outliers, we removed the top and bottom 2.5% of the wage and hours distributions. Also, following **?**, who also used the NLSY79 data, we only used assets that satisfy the following fomula:

$$-2500 \times (t-10) \leq A(t) \leq 10{,}000 \times (t-15)$$

[9] The failure to treat school attendance as a choice variable potentially creates two types of biases. Suppose that once people leave school they rarely return. If people decide when to leave school based on the wage draws they receive (i.e., the shock to human capital $\varepsilon_{1,t+1}$ in Equation (4)), then people will tend to have relatively high levels of wages (human capital) in the first period after leaving school. If the wage process exhibits any subsequent mean reversion, this may lead to understatement of the gradient of wages with respect to experience in the early postschool years. This would, in turn, cause us to underestimate the return to human capital investment in the early postschool years, which would cause us to underestimate the i.e.s. Suppose on the other hand, that people often return to school in periods when they receive very low wage draws. Failure to account for this is analogous to ignoring corner solutions in labor supply, which we have argued would be likely to bias estimates of the i.e.s. toward zero. In any case, omitting the schooling choice, as well as omitting the choice of working zero hours, may result in bias in our estimated parameters. We left those choices out of the model because (1) it imposes even more computational burden in the estimation routine, and (2) most of the labor supply literature, such as **?**, **?**, and **?** focus exclusively on workers with positive hours and omit schooling choices from their models as well. But estimating intertemporal labor supply models with corner solutions and schooling choices would be a promising future line of research. **?** estimate such a model, but they discretize the hours choices.

Table 1
MEAN AGE PROFILES

Age	Hourly Wage	Hours	Total Wealth
20	5.785 (2763)	1531.7 (2837)	6334.8 (207)
21	5.998 (3220)	1578.6 (3348)	6178.5 (568)
22	6.697 (3396)	1725.7 (3536)	8933 (956)
23	7.12 (3420)	1866.7 (3557)	9163 (1361)
24	7.342 (3308)	1989.4 (3480)	2889.9 (1751)
25	7.99 (3187)	2042 (3360)	-9276.6 (1885)
26	9.294 (2987)	2101.6 (3160)	1205.6 (2189)
27	9.098 (2871)	2134.7 (3048)	4651.3 (2504)
28	10.1 (2785)	2182.1 (2976)	-29888 (2538)
29	9.426 (2499)	2200.7 (2670)	-37101 (2235)
30	12.36 (2028)	2224.3 (2188)	-25726 (1903)
31	15.24 (1649)	2243.9 (1785)	-12168 (1614)
32	13.6 (1206)	2238.1 (1307)	19201 (1313)
33	22.98 (851)	2253.1 (922)	27379 (1098)
34	11.39 (554)	2246.7 (608)	55264 (649)
35	11.57 (291)	2294.2 (325)	84001 (299)
36	10.01 (65)	2283.5 (71)	58172 (67)

NOTE: Sample sizes are in parentheses

Table 2
QUANTILE AGE PROFILES

	Wage Quantiles				Hours Quantiles		
Age	25%	50%	75%		25%	50%	75%
20	3.26	4.752	6.641	800	1640	2080	
21	3.509	5.138	7.188	892	1694	2095	
22	3.912	5.645	7.98	1127	1948	2185	
23	4.14	6.096	8.623	1402	2080	2294	
24	4.527	6.533	9.203	1640	2080	2357	
25	4.739	7	9.833	1760	2080	2395	
26	5.079	7.404	10.38	1880	2080	2438	
27	5.389	7.754	10.84	1925	2080	2486	
28	5.633	8.142	11.35	2000	2120	2531	
29	5.673	8.318	11.65	2016	2145	2560	
30	5.851	8.611	12.02	2060	2162	2580	
31	5.989	8.76	12.43	2051	2170	2600	
32	6.077	8.983	12.98	2068	2134	2600	
33	6.166	8.917	12.66	2080	2165	2600	
34	6.737	9.378	13.24	2080	2161	2600	
35	6.581	9.316	13.72	2080	2200	2619	
36	6.388	10.18	12.37	2080	2151	2667	

where t is the age of the individual. This was necessary because there were some assets whose values were either extremely high or low. After censoring the data, the sample means are closer to the medians (see Table 3). In the estimation, we treat outliers as missing values.

In the data, the percentage of the individuals with zero hours supplied is 17% at age 20, falls below 10% at age 25, and is 9.6% overall. Hence, for the estimation of this article, we adopt the conventional approach of the prime age male labor supply literature and assume only interior solutions. We start estimating the model on individuals after they finish schooling and when they start working positive hours, and we exclude individuals who have a year of zero hours of work after that. The estimation of the intertemporal labor supply model with corner solutions using the dynamic programming ML approach is left for future research. Because of the computational burden of the dynamic programming and estimation routine, we restrict the heterogeneity to be only on the observed educational outcome. That is, we divide individuals into four educational types: type 1 individuals are those who did not graduate from high school (final grade less than 12 years), type 2 are those who graduated from high school (final grade equal to 12 years), type 3 are those who attended some college (final grade less than 16 years), type 4 are those who graduated from college (final grade more than or equal to 16 years).[10]

[10] While we do not include unobserved heterogeneity in the model, persistence in wages is captured by the persistent nature of the shocks to human capital.

Table 3
MEAN AGE PROFILES AFTER CENSORING

Age	Mean Wage	Sample Size	Mean Hours	Sample Size	Zero Hours	Mean Assets	Sample Size
20	5.147	2624	1567.5	2769	17.14	3478.9	202
21	5.582	3059	1616.3	3266	14.94	4633.2	561
22	6.09	3227	1766.5	3451	13.63	5404.6	931
23	6.603	3249	1910.9	3470	11.74	6676.6	1337
24	7.07	3143	2034.3	3398	10.24	8501.6	1706
25	7.5	3028	2089.9	3277	8.37	10333	1831
26	7.965	2837	2146.1	3088	6.81	11195	2128
27	8.34	2728	2183	2973	5.95	12312	2434
28	8.748	2646	2231.2	2903	5.97	13818	2439
29	8.943	2374	2250.5	2605	5.42	15733	2143
30	9.349	1927	2272.1	2135	6.09	14695	1818
31	9.656	1566	2289.9	1744	5.1	14431	1530
32	10.04	1145	2284.6	1276	6.24	13029	1265
33	9.866	808	2300.3	900	6.49	14730	1060
34	10.51	527	2292.1	594	7.03	24411	604
35	10.81	276	2337	318	8.19	32658	252
36	9.768	62	2283.5	71	7.79	22175	59

The total sample of white males who are at least 20, have completed schooling, are not in the military, and whose schooling record is known is 4418 individuals. Among them, 718 individuals are high school dropouts, 1980 individuals are high school graduates, 869 individuals attended some college, and 851 individuals are college graduates. Among them, we only use the individuals who completed schooling before age 25 and whose wage and hours data are available for at least 6 years. The total sample size of those left is 2143 individuals. Then, we remove those people that have zero hours at some point after the starting age. After that, the remaining sample is 1972 individuals.

Finally, we restricted the sample size for the estimation by randomly choosing 1000 people out of 1972 individuals. The total number of person year observations is 7465. The total number of wage observations is 7465, that of the hours observations is 7465, and that of the asset observations is 4323. Notice that people with missing data for several periods are still carried forward (as described in the appendix).

6. ESTIMATION RESULTS AND SIMULATION EXPERIMENTS

6.1. *Main Estimation Results.* We report the parameter estimates in Table 4. The key result is the estimate of the disutility of labor parameter, which is 1.262. The implied elasticity of intertemporal substitution is

$$i.e.s. \equiv b_2 \equiv \frac{1}{a_2 - 1} = 3.820$$

Table 4

ESTIMATION RESULTS

Utility Function Parameters			
Disutility of labor			
	a_2	Disutility of labor curvature	$1.2618(8.504 \times 10^{-4})$
b	b_n	Non-high school	$1.831 \times 10^{-5}(1.89 \times 10^{-7})$
	b_h	High school graduate	$1.65 \times 10^{-5}(6.80 \times 10^{-8})$
	b_c	Some college	$1.62 \times 10^{-5}(1.04 \times 10^{-7})$
y	b_{cg}	College graduate	$1.75 \times 10^{-5}(2.04 \times 10^{-7})$
	σ_1	Std. error of disutility shock	$0.01156(6.75 \times 10^{-4})$
Consumption utility			
	a_1	Consumption CRRA	$0.2617(5.73 \times 10^{-4})$
	C_0	Constant	$0.017(5.97 \times 10^{-5})$
C_1	C_{1n}	Non-high school	$0.5859(-0.0108)$
	C_{1h}	High school graduate	$0.5241(-0.003821)$
	C_{1c}	Some college	$0.5175(-0.01022)$
	C_{1cg}	College graduate	$0.546(-0.01967)$
C_2	C_{2n}	Non-high school	$0.2259(-0.005984)$
	C_{2h}	High school graduate	$0.1672(-0.001954)$
	C_{2c}	Some college	$0.1294(-0.007231)$
	C_{2cg}	College graduate	$0.1517(-0.00627)$
	β	Discount factor	$0.9529(2.47 \times 10^{-4})$
Production Function Parameters [1]			
δ	δ_n	Non-high school	$0.404(0.002633)$
	δ_h	High school graduate	$0.3458(9.71 \times 10^{-4})$
	δ_c	Some college	$0.3189(0.002413)$
	δ_{cg}	College graduate	$0.3434(0.002145)$
k_0	k_{0n}	Non-high school	$0.01588(0.002521)$
	k_{0h}	High school graduate	$0.02843(0.002224)$
	k_{0c}	Some college	$0.05387(0.001278)$
	k_{0cg}	College graduate	$0.05719(0.002262)$
A_0	A_{0n}	Non-high school	$0.1304(6.91 \times 10^{-4})$
	A_{0h}	High school graduate	$0.1513(3.15 \times 10^{-4})$
	A_{0c}	Some college	$0.1536(6.59 \times 10^{-4})$
	A_{0cg}	College graduate	$0.1463(5.12 \times 10^{-4})$
A_1	A_{1n}	Non-high school	$-0.002139(1.83 \times 10^{-5})$
	A_{1h}	High school graduate	$-0.00342(1.62 \times 10^{-5})$
	A_{1c}	Some college	$-0.002915(7.11 \times 10^{-5})$
	A_{1cg}	College graduate	$-0.003329(6.94 \times 10^{-5})$
α	α_n	Non-high school	$0.2279(4.14 \times 10^{-4})$
	α_h	High school graduate	$0.2243(1.36 \times 10^{-4})$
	α_n	Some college	$0.2258(3.31 \times 10^{-4})$
	α_h	College graduate	$0.2275(3.31 \times 10^{-4})$
	B_2	$-B_2(h+d_1)$	$4.05 \times 10^{-4}(7.29 \times 10^{-7})$

B_1	Additive constant in capital term $B_1 + K$	$0.04021(7.29 \times 10^{-4})$
σ_0	Std. error of wage shock	$0.05781(6.05 \times 10^{-4})$
d_1	Additive constant in hours term $h + d_1$	367.2(6.035)
Mean Initial Assets		
$\bar{A}$	Mean initial assets when the starting age is 20	3250.8(458.6)
$\bar{A}$	Mean initial assets when the starting age is after 20	7190.4(631.1)
$V_{\bar{A}}$	Std. error, initial assets	2218.7(241.3)
Measurement Error Parameters		
$\sigma_{\xi 0}$	Initial period wage [2]	0.4909 (0.003626)
$\sigma_{\xi 1}$	Wage [3]	0.4643 (0.001333)
$\sigma_{\xi 2}$	Hours [4]	590.7 (2.156)
$\sigma_{\xi 31}$	Asset [5]	2623.5 (178.5)
$\sigma_{\xi 32}$	Asset	948.8 (11.98)

NOTES: Standard errors are in parentheses.

[a] $g(K,h,t) = A_0(1 + A_1(t-19))(B_1 + K)[(h+d_1)^{\alpha} - B_2(h+d_1)] + \delta K + k_0.$

[b] $K_{t_0}^D = K_{t_0}\xi_0, \quad \ln(\xi_0) \sim N\left(0, \sigma_{\xi,0}\right)$

[c] $K_t^D = K_t h_t \frac{\xi_{1,t}}{h_t^D}, \quad \ln(\xi_{1,t}) \sim N\left(0, \sigma_{\xi,1}\right)$

[d] $h_t^D = h_t + \xi_{2,t}, \quad \xi_{2,t} \sim N\left(0, \sigma_{\xi,2}\right)$

[e] $A_t^D = A_t + \xi_{3,t}, \quad \xi_{3,t} \sim N\left(0, \sigma_{\xi,3}\right), \quad \sigma_{\xi,3} = \sigma_{\xi,3,1} + \sigma_{\xi,3,2}(t-19).$

This elasticity estimate is reasonably close to the elasticity parameter macroeconomists typically use to calibrate real business cycle models (e.g., **?**, obtain an elasticity estimate that is around 5, and **?**, uses 2 in his calibration exercise).

Another key parameter of interest is a_1, the CRRA parameter, which we estimate to be 0.262. Thus, the intertemporal elasticity of substitution in consumption (i.e.s.-c) is

$$\frac{1}{a_1 - 1} = -1.354$$

This is quite different from conventional estimates. Typically, a_1 is estimated to be around -2, which implies that the i.e.s.-c is around -1/3 (see **?**). **?** also estimate a dynamic model of labor supply, human capital accumulation, and saving, and they obtain an estimate of the i.e.s.-c, which is around -2 (i.e., $a_1 \approx 0.5$), which is more similar to our estimate of i.e.s.-c = -1.354 than most of the prior literature. **?** (p. 1078) argue that prior estimates of this i.e.s.-c had to imply a high degree of prudence (2 - a_1) in order to rationalize the fact that youths with steep age-earnings profiles do not borrow substantially against future income. They argue that their estimate of a_1 is larger than the values of about -2 typically estimated in the literature because their model incorporates borrowing constraints. Similarly, we use age effects in the marginal utility of consumption to help explain the failure of youths to borrow against future income. This may account for our larger estimate of a_1. Recently, **?** have argued that experimental evidence is consistent with $a_1 \approx 0.50$.

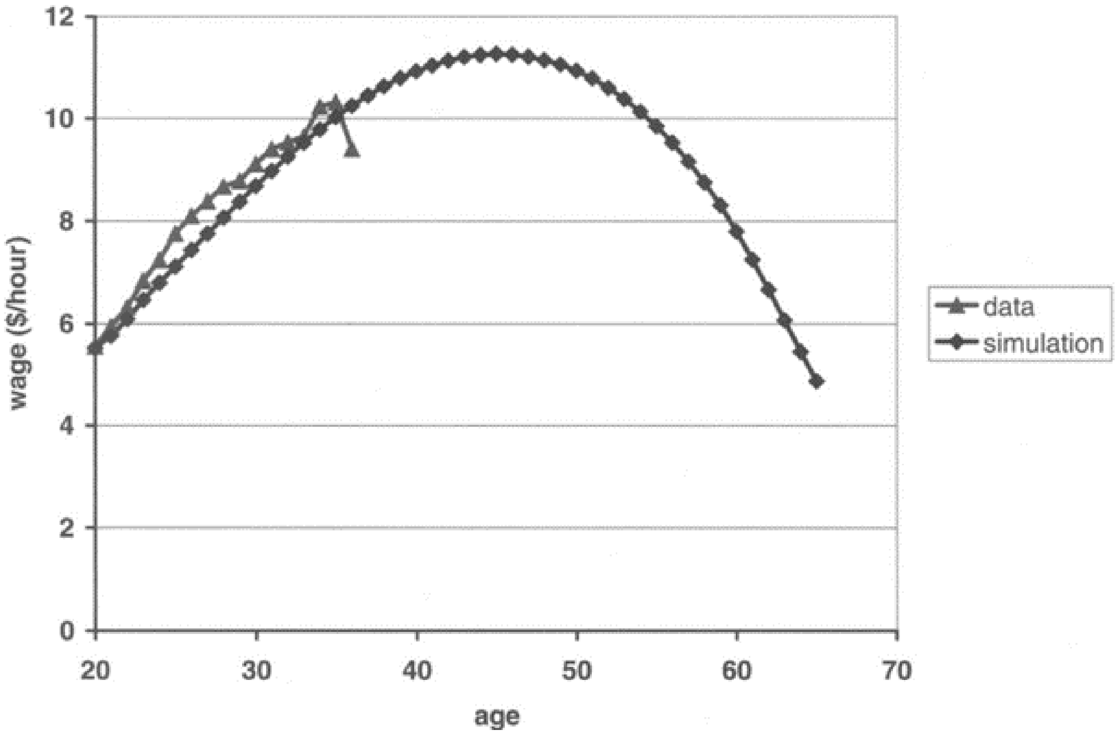


Figure 3: AGE-WAGE PROFILES

6.2. *Model Fit.* To evaluate the fit of the model, we artificially generated 1000 individual life-cycle paths from ages 20 to 65 using the estimated parameters. The various age profiles in the simulated data are reported in Table 5. Figures 3-5 compare simulated age profiles of wages, labor supply, and assets with those of the data. The simulated profiles resemble the actual profiles reasonably well. Notice that the age-hours profile before retirement is rather flat compared to the significant humped shape of the simulated age-wage profile. The model is able to reconcile this fact with a large i.e.s. because of the human capital effect, as we have discussed earlier.

The simulated asset paths indicate that, even though labor income is small when people are young, individuals do not go into sizeable debt early in their life. In a model with perfect capital markets, this implies that, for the same consumption level, the value of consumption is smaller when people are young. As noted earlier, alternative explanations could be the existence of some finance constraints (see **?**) or some nonseparability between consumption and labor. We did not pursue these explanations in this article. The out-of-sample predictions of the model for wages, hours, and assets look quite reasonable. The simulated age-hours path predicts retirement behavior of the agents at older ages, although that occurs somewhat sooner in the simulation than has been the case in the United States. On the other hand, later cohorts have been retiring at younger ages, so we cannot rule out that this prediction may be accurate for the cohort used here. It is also worthwhile to notice that the model predicts large asset accumulation between ages 40 and 60, and dissaving afterwards, which is the actual pattern of savings and dissavings (see **?**).

Table 5

SIMULATED MEAN AGE PROFILES

Age	Mean Wage	Mean Hours	Mean Assets	$[(1+r)\beta]^{t-20}u_C(t)$
20	5.517	1899.4	3239.2	1.26×10^{-5}
21	5.754	1943.0	4031.4	1.27×10^{-5}
22	6.088	1993.7	4614.1	1.26×10^{-5}
23	6.464	2040.5	5942.4	1.25×10^{-5}
24	6.799	2087.6	6784.4	1.24×10^{-5}
25	7.114	2132.9	7977.7	1.24×10^{-5}
26	7.437	2176.8	9182.3	1.24×10^{-5}
27	7.754	2219.3	10,389	1.24×10^{-5}
28	8.066	2260.5	11,582	1.24×10^{-5}
29	8.378	2300.1	12,744	1.24×10^{-5}
30	8.685	2337.3	13,842	1.24×10^{-5}
31	8.974	2372.9	14,834	1.24×10^{-5}
32	9.256	2406.6	15,674	1.25×10^{-5}
33	9.526	2437.8	16,300	1.25×10^{-5}
34	9.784	2466.4	17,820	1.25×10^{-5}
35	10.03	2492.0	20,230	1.25×10^{-5}
36	10.25	2515.1	23,517	1.25×10^{-5}
37	10.45	2534.7	27,656	1.25×10^{-5}
38	10.63	2551.0	32,615	1.25×10^{-5}
39	10.79	2562.3	38,344	1.25×10^{-5}
40	10.92	2569.5	44,772	1.25×10^{-5}
41	11.04	2572.6	51,828	1.25×10^{-5}
42	11.14	2569.8	59,432	1.25×10^{-5}
43	11.20	2560.2	67,494	1.25×10^{-5}
44	11.24	2545.0	75,890	1.25×10^{-5}
45	11.27	2522.1	84,519	1.25×10^{-5}
46	11.25	2491.4	93,256	1.25×10^{-5}
47	11.21	2452.0	101,973	1.25×10^{-5}
48	11.14	2402.2	110,531	1.25×10^{-5}
49	11.06	2342.0	118,779	1.25×10^{-5}
50	10.94	2271.1	126,622	1.25×10^{-5}
51	10.79	2189.6	133,972	1.26×10^{-5}
52	10.60	2096.7	140,751	1.26×10^{-5}
53	10.38	1990.8	146,866	1.27×10^{-5}
54	10.13	1872.9	152,228	1.27×10^{-5}
55	9.849	1742.5	156,759	1.28×10^{-5}
56	9.530	1597.5	160,371	1.30×10^{-5}
57	9.161	1441.1	162,959	1.31×10^{-5}
58	8.755	1274.1	164,440	1.32×10^{-5}
59	8.309	1102.0	164,709	1.33×10^{-5}
60	7.799	933.21	163,702	1.34×10^{-5}
61	7.250	777.70	161,383	1.35×10^{-5}
62	6.669	633.47	157,795	1.36×10^{-5}
63	6.070	488.99	153,096	1.38×10^{-5}
64	5.452	434.42	147,231	1.45×10^{-5}
65	4.877	251.58	141,247	1.36×10^{-5}

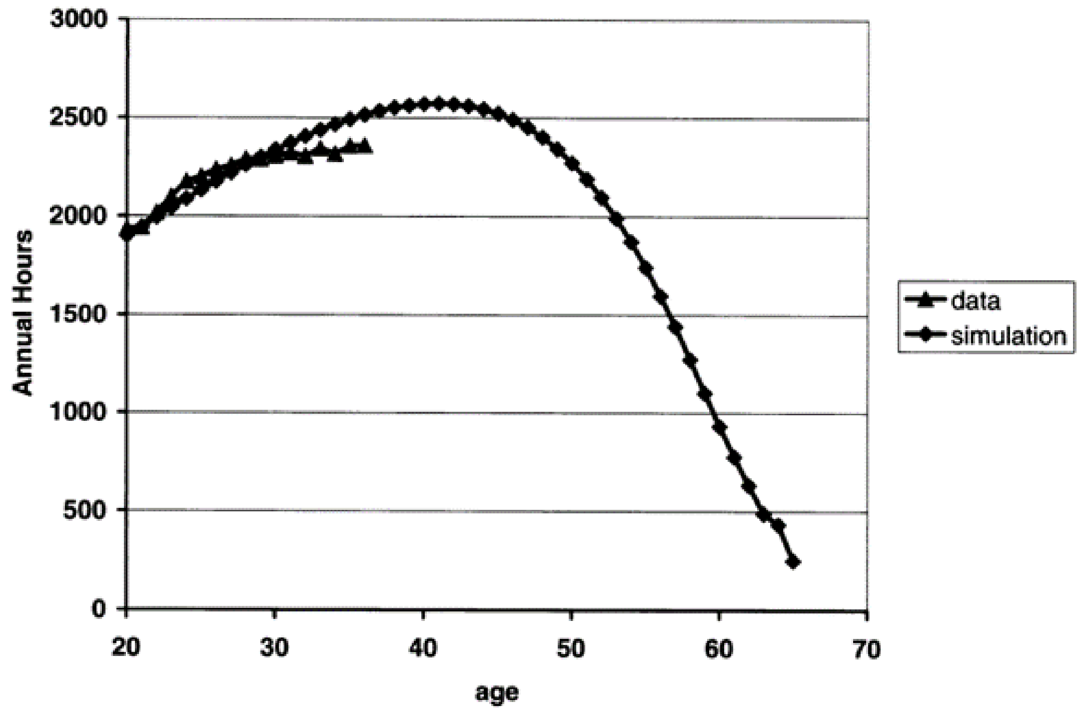


Figure 4: AGE-HOURS PROFILES

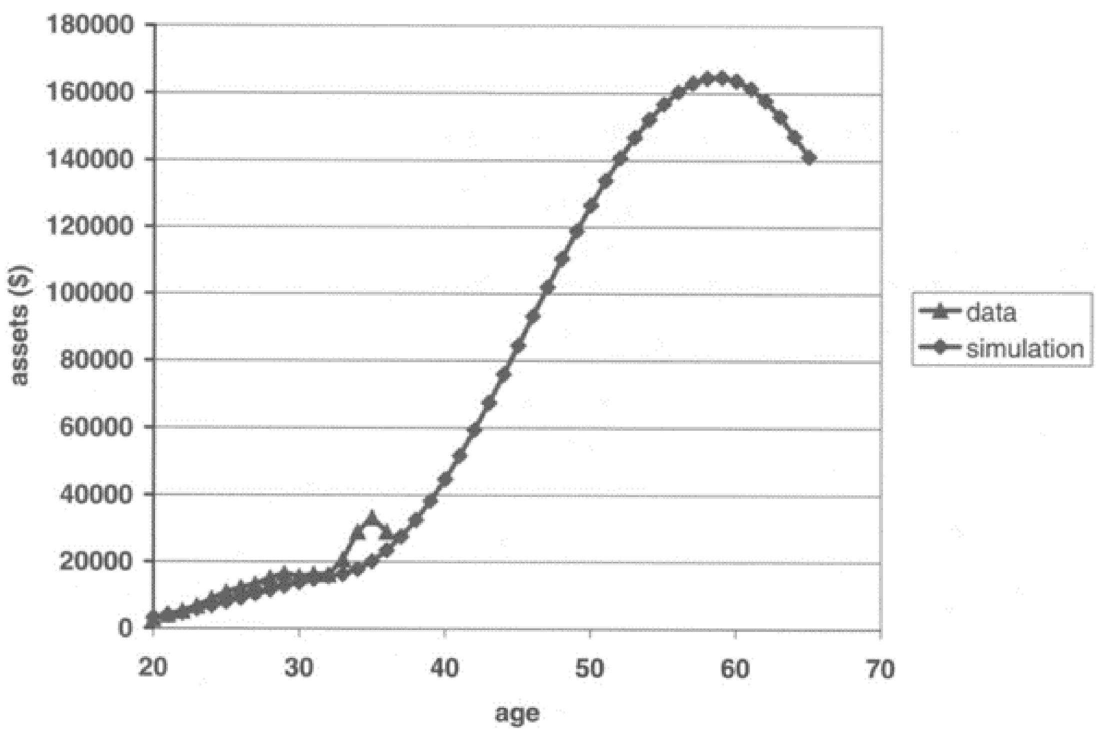


Figure 5: AGE-ASSET PROFILES

The performance of the approximate DP solution can be indirectly inferred from the age profile of the discounted marginal utility of consumption. In Table 5, we report the simulated mean age profile of $[(1.0+r)\beta]^{t-20}u_C(Ct,t)$, which should be constant over age.[11] The profile is roughly constant, except it rises a bit in the late 50s and 60s. This gives indirect evidence that the solution algorithm seems to work fairly well overall.

6.3. *Using Simulation Exercises to Assess the Bias in Conventional Elasticity Estimates.* Using the simulated data, we conduct OLS and IV exercises to estimate the elasticity of intertemporal substitution using the methods of MaCurdy and Altonji. That is, we estimate the following equation via OLS and IV:

$$\Delta\ln(h_t) = \text{Const} + b_2\Delta\ln(W_t) + \zeta_t$$

where ζ_t is an error term. From the values of estimated coefficient b_2, we also recover the disutility of labor parameter as follows:

$$a_2 = \frac{1}{b_2} + 1$$

In Table 6, we report the results of estimation on simulated data, and in Table 7, we report the results of estimation on the simulated data that we cleaned by using an outlier elimination procedure that is similar to **?**.[12] The instruments for our IV exercise include a constant term, experience, experience squared, and the twice lagged wage. All the OLS and IV results are the average of 10 repetitions with independently simulated data. We used various age groups for the exercise, starting with simulation from age 20 to 64. Then, in Table 8 we obtained results with ages from 20 to 56, 20 to 46, and 20 to 36, which is the age group in the NLSY79.

Note that in Tables 6 and 7, the elasticity estimates obtained from the simulated data using conventional methods are low, compared to the true value used to generate the data (3.820). This is true regardless of whether we remove outliers or not, which indicates that the downward bias does not depend much on the outliers. That is, the true i.e.s. in the simulated data is much higher than the estimates obtained using conventional IV methods. These results confirm the point that we get biased (toward 0) estimates of the i.e.s. if we do not explicitly allow for human capital accumulation in the model.

Also note that the OLS estimates of the intertemporal elasticity b_2 are much smaller than the IV estimates. The theory implies that in the OLS case, there is another reason for downward bias

[11]The model allows the value that agents place on consumption to vary with age, as determined by the $A(t)$ term in $U(c_t,t)$. According to our estimates, for a high school graduate, $A(t)$ is equal to C_0 times 0.524 (=C_1) at age 20, it rises to C_0 times 0.6913 (=C_1+C_2) at age 25, and then rises to C_0 times 1.0 at age 33 and above. Results for other education categories are similar. Thus, at younger ages, less consumption is needed to reduce the marginal utility of consumption to any given level.

[12]The outlier elimination rules are:

1. Annual hours worked must be less than 4680 hours.
2. For the calculation of changes in log earnings, the absolute value of the difference in a person's average hourly earnings in adjacent years cannot exceed \$16 or a change of 200%.
3. For the calculation of the changes in log hours, the absolute value of the difference in the annual hours of work in adjacent years cannot exceed 3000 hours or a change of 190%.

Table 6

ML, OLS, IV RESULTS

Age		b_2	a_2
20 – 36	ML	3.82(0.0124)	1.26(0.000850)
Simulated data			
20 – 64	OLS	−0.444(0.00248)	−1.25(0.0126)
	3SLS	0.971(0.0941)	2.03(0.0999)
20 – 56	OLS	−0.301(0.00247)	−2.33(0.0273)
	3SLS	1.13(0.259)	1.89(0.204)
20 – 46	OLS	−0.270(0.00280)	−2.70(0.0382)
	3SLS	0.537(0.217)	2.86(0.753)
20 – 36	OLS	−0.293(0.00366)	−2.41(0.0426)
	3SLS	0.325(0.256)	4.08(2.42)
NLSY79 data			
20 – 36	OLS	−0.231(0.00659)	−3.33(0.124)
	3SLS	0.260(0.0769)	4.85(1.14)

NOTE: Delta method is used to calculate the standard errors for b_2 for ML and a_2 for OLS and 3SLS results.
Std. errors are in parentheses. Instruments: const, experience (which is age-19), experience squared, twice lagged wage.

TABLE 7

OLS, IV RESULTS WITH CLEANED DATA

Age		b_2	a_2
20–36	ML	3.82 (0.0124)	1.26 (0.000850)
Simulated data			
20–64	OLS	−0.168 (0.0299)	−4.94 (1.056)
	3SLS	1.21 (0.196)	1.83 (0.134)
20–56	OLS	−0.171 (0.0895)	−4.85 (3.067)
	3SLS	1.68 (0.569)	1.60 (0.202)
20–46	OLS	−0.171 (0.00590)	−4.87 (0.203)
	3SLS	0.655 (0.521)	2.53 (1.22)
20–36	OLS	−0.177 (0.00775)	−4.65 (0.248)
	3SLS	0.0750 (0.576)	14.3 (102)
NLSY79 data			
20–36	OLS	−0.232 (0.00607)	−3.31 (0.113)
	3SLS	0.142 (0.0731)	8.03 (3.61)

NOTE: Delta method is used to calculate the standard errors for b_2 for ML and a_2 for OLS and 3SLS results.
Std. errors are in parentheses. Instruments: const, experience (which is age-19), experience squared, twice lagged wage.

Table 7 OLS, IV RESULTS WITH CLEANED DATA

TABLE 8

ML, OLS, IV RESULTS OF ALTONJI ESTIMATION

Age		b_2	a_2
20–36	ML	3.82 (0.0124)	1.26 (0.000850)
Simulated data			
20–64	OLS	−0.206 (0.00459)	−3.85 (0.108)
	3SLS	2.81 (0.0804)	1.36 (0.0102)
20–56	OLS	−0.272 (0.00274)	−2.68 (0.0371)
	3SLS	−0.218 (0.0306)	−3.59 (0.646)
20–46	OLS	−0.246 (0.00299)	−3.07 (0.0494)
	3SLS	0.322 (0.0467)	4.11 (0.451)
20–36	OLS	−0.281 (0.00380)	−2.56 (0.0481)
	3SLS	0.476 (0.182)	3.10 (0.803)

NOTE: Delta method is used to calculate the standard errors for b_2 for ML and a_2 for OLS and 3SLS results.
Std. errors are in parentheses. Instruments: const, experience (which is age-19), experience squared, experience cubed.

Table 8 ML, OLS, IV RESULTS OF ALTONJI ESTIMATION because, in this case, the error term is correlated with the regressor due to the income effect. This is confirmed in our numerical example.

It is also interesting to examine how the IV estimates vary with the age composition of the simulated data. It seems that if older individuals are heavily represented in the data, then the elasticity estimates tend to increase, with the maximum elasticity estimates being on average those obtained using the 20-56 age group. In particular, for that age group, we obtain 1.13 using the raw simulated data, and 1.68 using the simulated data without outliers. For younger age groups, the elasticity estimates are much lower, for example, for the 20-36 age group using all the simulated data, we obtain 0.325, which drops to 0.0750 when outliers are removed. Those results underscore the fact that the human capital component of the return to labor supply is much greater for the young.

We also report (in the bottom panel of Tables 6 and 7) the IV results obtained from the NLSY79 data that were used to obtain the ML estimates. The elasticity estimates are: 0.260 for the original NLSY79 data and 0.142 for the cleaned data. These estimates of the i.e.s. are much smaller than the one derived from the ML estimation, again suggesting that failure to account for human capital accumulation leads to downward bias.

6.4. *More on Model Fit.* A comparison of OLS estimation results on the NLSY79 and on the simulated data is another method of assessing the fit of the model. Notice that for the age group 20-36, the OLS estimates of a regression of the log hours change on the log wage change produces similar results for the simulated and NLSY79 data. The OLS estimate from the simulated data is -0.293 whereas that from the NLSY79 data is -0.231 (see Table 6, bottom 2 panels). Thus, the model seems to reproduce not only the average age profiles of the data, but also the negative raw correlation between log wage changes and log hours changes observed in the data.

It is also interesting to compare the persistence of wages, hours, and assets in the simulated versus the actual data. In Table 9, we report results where we regress the log wage, log hours, and assets from the simulated and NLSY79 data on a constant term, lagged values and age variables. For both the simulated and NLSY79 data, we removed outliers. The results show that the estimated model captures well the persistence of the log wages and assets in the data. For the regression using the simulated wage data, the coefficient of the lagged log wage is 0.836 if age variables are not included, and 0.823 if the age variables are included. In the regression using the NLSY79 data, these coefficients are 0.879 and 0.878, respectively. In the regression using simulated assets, the coefficient on lagged assets is 0.591 if the age variables are not included, and 0.551 if the age variables are included. In the regression using the NLSY79 data, these coefficients are 0.662 and 0.616, respectively.

On the other hand, in the hours regressions, there are some discrepancies in the persistence coefficients between the simulated and NLSY79 data. For the regression using simulated hours data, the persistence coefficients are 0.413 without age effects and 0.362 with the age effects. These coefficients are much lower than those using the NLSY data, i.e., 0.708 and 0.705, respectively. One potential reason for the discrepancy may be the assumption made in the model that the measurement error of hours is i.i.d. normally distributed, which may have been overly simplistic. In order to examine the effect of this assumption, we report results of another regression using the simulated data of log hours that does not include any measurement error. In this regression, the coefficient of the lagged log hours is 0.941 without age effects and 0.939 with age effects. These results imply that it is the measurement error that reduces the persistence of the simulated log

TABLE 9

WAGE, HOURS, AND ASSET REGRESSIONS

Simulated Data		
Dependent Variable	Log Wage	Log Wage
Intercept	0.309 (0.0117)	0.210 (0.0182)
Age	—	0.00137 (0.00438)
Age squared	—	0.000103 (0.000234)
Lagged log wage	0.836 (0.00594)	0.823 (0.00616)
Standard error	0.357	0.356
R^2	0.700	0.702
F-statistic	19,817	6668
Sample size	8515	8515
Dependent Variable	Asset	Asset
Intercept	5507 (145)	2685 (5642)
Age	—	−345 (410)
Age squared	—	16.1 (7.29)
Lagged asset	0.591 (0.00736)	0.551 (0.00758)
Standard error	14,978	14,799
R^2	0.302	0.318
F-statistic	6452	2325
Sample size	14,949	14,949
Dependent Variable	Log Hours[a]	Log Hours[a]
Intercept	4.53 (0.0740)	4.34 (0.139)
Age	—	0.0312 (0.00876)
Age squared	—	−0.000364 (0.000155)
Lagged log hours	0.413 (0.00962)	0.362 (0.00992)
Standard error	0.241	0.236
R^2	0.178	0.206
F-statistic	1846	735
Sample size	8515	8515
Dependent Variable	Log Hours[b]	Log Hours[b]
Intercept	0.476 (0.0158)	0.459 (0.0221)
Age	—	0.00201 (0.000878)
Age squared	—	$-3.51 \times 10^{-5}(1.55 \times 10^{-5})$
Lagged log hours	0.941 (0.00187)	0.939 (0.00279)
Standard error	0.0261	0.0261
R^2	0.953	0.953
F-Statistic	210,082	70,051
Sample size	10,462	10,462

TABLE 9

CONTINUED

NLSY79 Data		
Dependent Variable	Log Wage	Log Wage
Intercept	0.274 (0.00803)	0.263 (0.00955)
Age	—	0.00449 (0.00208)
Age squared	—	−0.000285 (0.000138)
Lagged log wage	0.879 (0.00393)	0.878 (0.00414)
Standard error	0.227	0.227
R^2	0.769	0.769
F-statistic	50,134	16,716
Sample size	15,030	15,030
Dependent Variable	Asset	Asset
Intercept	10,206 (284.8)	387 (1227)
Age	—	734 (321)
Age squared	—	66.4 (19.2)
Lagged asset	0.662 (10.9)	0.616 (0.0126)
Standard error	21,312	20,526
R^2	0.254	0.308
F-statistic	2625	1145
Sample size	7732	7732
Dependent Variable	Log Hours	Log Hours
Intercept	2.26 (0.0428)	2.27 (0.0430)
Age	—	0.00277 (0.00151)
Age squared	—	$-1.27 \times 10^{-4} (1.01 \times 10^{-4})$
Lagged log hours	0.708 (0.00556)	0.705 (0.00565)
Standard error	0.166	0.166
R^2	0.519	0.519
F-Statistic	16,217	5410
Sample size	15,030	15,030

NOTE: Std. errors are in parenthesis.
[a]Simulated hours include measurement error.
[b]Simulated hours do not include measurement error.

Table 10 CONTINUED

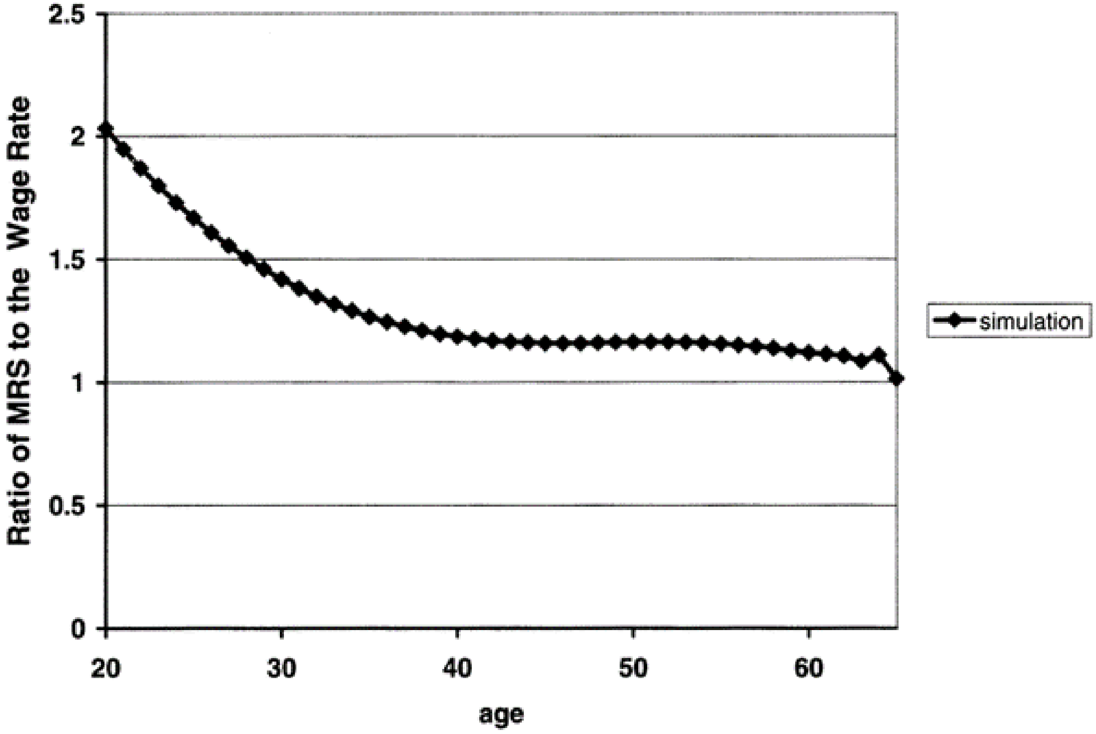


Figure 6: SIMULATED MRS/WAGE PROFILES

hours in the original regression. An estimation exercise using a model that allows for a richer specification of measurement error is left for future research.

6.5. *How Human Capital Accumulation Affects Estimated and Actual Labor Supply Elasticities* To assess the importance of human capital accumulation for the labor supply decision, we also report in Table 11 the age profile of the mean marginal rate of substitution between consumption and labor supply, and the marginal rate of substitution divided by the wage. The latter is also shown in Figure 6. Note that the marginal rate of substitution is significantly higher than the real wage early in life. At age 20 it is 2.0 times greater than the real wage. Then, the marginal rate of substitution becomes closer to the actual wage rate at later stages of the career. The bias in the MaCurdy and Altonji estimation method arises from the fact that they do not recover the marginal rate of substitution, or the "effective wage," which is higher than the observed wage when there is human capital accumulation.

This brings us to an important point, which is that IV methods cannot be used to solve the problem of bias in estimates of the substitution elasticity created by human capital accumulation. The reason that standard IV results are biased can best be described as follows. Define W_t as the marginal rate of substitution between labor supply and consumption in period t. Also, define

$$\eta_t = \ln(\tilde{W}_t) - \ln(W_t)$$

Then, from the definition of the marginal rate of substitution,

$$\ln(\tilde{W}_t) = \ln b - \ln A(t) + (a_2 - 1)\ln h_t - (a_1 - 1)\ln C_t + \ln\varepsilon_{2,t}$$

TABLE 10

SIMULATED MRS OVER THE LIFE CYCLE

Age	MRS	MRS/Wage	Age	MRS	MRS/Wage
20	11.22	2.032	43	13.01	1.164
21	11.22	1.948	44	13.03	1.162
22	11.39	1.871	45	13.02	1.159
23	11.63	1.799	46	13.01	1.158
24	11.76	1.730	47	12.97	1.159
25	11.86	1.668	48	12.92	1.161
26	11.95	1.608	49	12.84	1.162
27	12.03	1.554	50	12.71	1.164
28	12.12	1.506	51	12.54	1.163
29	12.21	1.461	52	12.32	1.164
30	12.28	1.419	53	12.06	1.163
31	12.36	1.382	54	11.75	1.160
32	12.43	1.348	55	11.39	1.157
33	12.51	1.318	56	10.97	1.151
34	12.58	1.291	57	10.49	1.146
35	12.65	1.266	58	9.968	1.139
36	12.71	1.245	59	9.371	1.128
37	12.76	1.226	60	8.732	1.120
38	12.82	1.211	61	8.078	1.115
39	12.86	1.196	62	7.373	1.107
40	12.90	1.186	63	6.583	1.086
41	12.95	1.178	64	6.044	1.110
42	12.98	1.169	65	4.943	1.015

Table 11 SIMULATED MRS OVER THE LIFE CYCLE

If we first difference away the consumption term, we get the following expression:

$$\begin{aligned}\Delta\ln(h_t) &= \text{Const} + \frac{1}{a_2-1}\Delta\ln(\tilde{W}_t) + \zeta_t \\ &= \text{Const} + \frac{1}{a_2-1}\Delta\ln(W_t) + \frac{1}{a_2-1}\Delta\eta_t + \zeta_t\end{aligned}$$

where ζ_t is a function of the error term u_t in the log linearized consumption Euler equation and of $\Delta\ln\varepsilon_{2,t}$:

$$\zeta_t = -\frac{1}{a_2-1}(u_t + \Delta\ln\varepsilon_{2,t}.$$

Note that, using equation (9), η_t can be expressed as follows:

$$\eta_t = \ln(\tilde{W}_t) - \ln(W_t) \approx \ln\frac{h_h\beta E\varepsilon_{1,t+1}V_{K,t+1}}{u_C W_t}$$

The problem is that conventional instruments for $\Delta\ln(W_t)$ are correlated with $\Delta\eta_t$. For example, age is correlated with the expected marginal value of human capital, $EV_{K,t+1}$. The older the individual, the closer he is to the retirement period, hence the less the number of possible future periods where human capital is used. Therefore, the older the individual, the less the marginal value of human capital. Table 11 shows how $\Delta\eta_t$ is negatively correlated with age, because the human capital effect decreases with age. In this case, the elasticity estimates from IV estimation using age as one of the instruments will likely be negatively biased. More generally, any variable that helps to predict wage growth ($\Delta\ln(W_t)$) is likely to be correlated with $V_K(t+1)$ and U_C.

To further illustrate how conventional methods of estimation are biased, we also report the results based on the **?** method of IV estimation, where consumption is used as a proxy for marginal utility of wealth. That is, we use the simulated data to estimate the equation

$$\ln(h_t) = \text{Const} + \frac{1}{a_2-1}\ln(W_t) + \frac{a_1-1}{a_2-1}\ln C_t - \frac{1}{a_2-1}\ln\varepsilon_{2,t}$$

As we did not simulate measurement error in consumption, we used the simulated consumption value without measurement error as the regressor, together with simulated wage with measurement error. The intertemporal elasticity is again estimated to be much lower than the true value, confirming our claim that omission of human capital effects biases the elasticity estimates downwards. In Figure 7, we plot the age-wage profile, which corresponds to the substitution effect, or the marginal return to labor supply from increasing wage income (which, in this case, is the real wage). We also plot the age-human capital effect profile, which is the real value of the marginal return to labor supply from increasing human capital (which, in this case, is $\beta E_t g_h \varepsilon_{1,t+1} V_{K,t+1}(A_{t+1}, K_{t+1}, \varepsilon_{2,t+1})/u_C(C_t,t)$). We can see that until the age 50, the increase in substitution effect cancels out with the human capital effect, resulting in their sum (which is the total shadow value of labor supply) being roughly flat. As discussed earlier, this explains the flat age-hours profile until the age 50. Afterwards, both the substitution effect and the human capital effect decrease as individuals start to retire.

Finally, we consider the implications of our model for the elasticity of hours with respect to wages. In Figure 8, we plot the change in age t hours due to a 2% temporary increase in the age t wage payment, holding human capital fixed. This experiment can be interpreted as a temporary

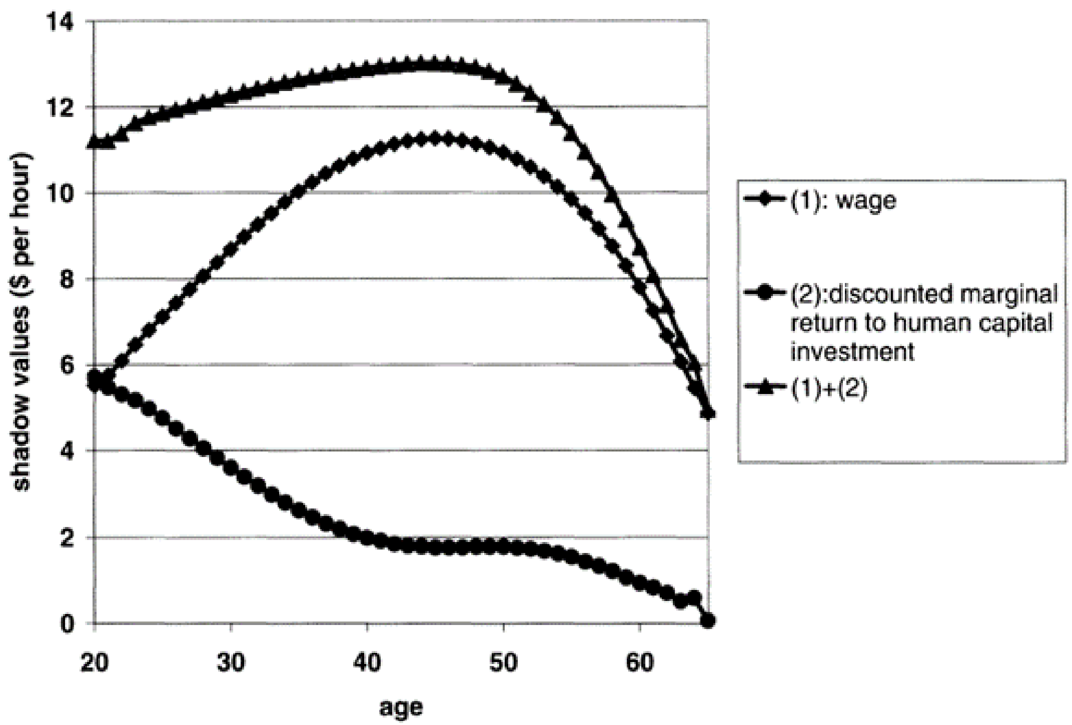


Figure 7: AGE SHADOW WAGE PROFILES

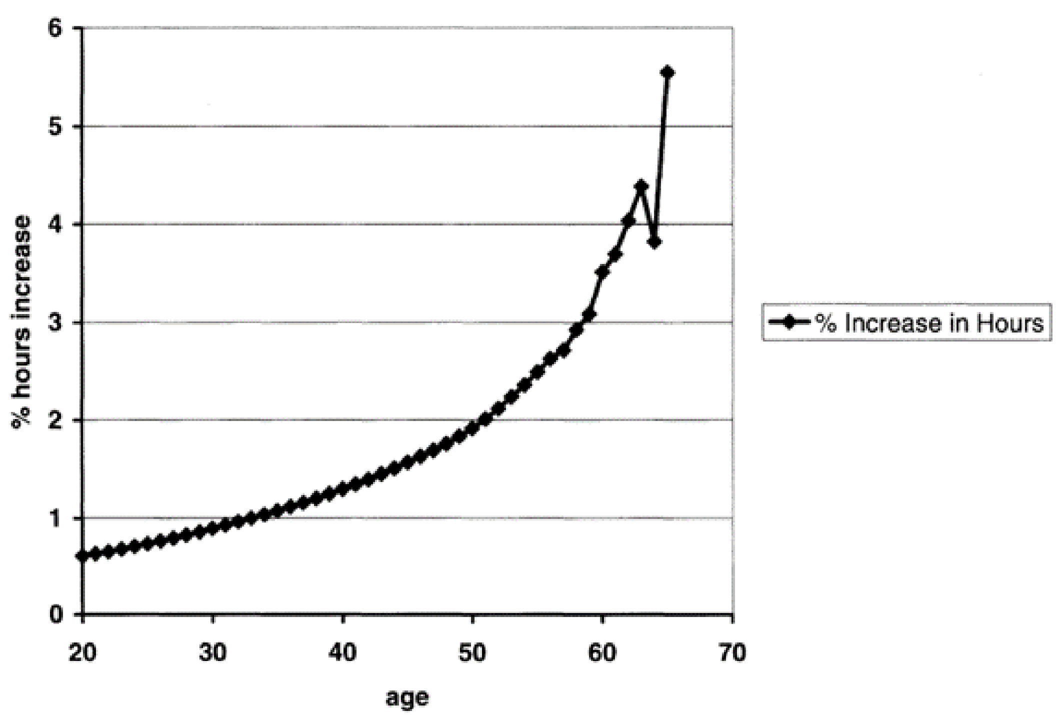


Figure 8: PERCENT INCREASE IN HOURS DUE TO TEMPORARY 2% INCREASE IN WAGES

increase in rental rate for human capital by 2%. For a person at age 20, it only increases hours by around 0.6%. But the elasticity of hours with respect to the current wage becomes larger with age. At around age 60, hours increase by more than 4%. That is, with age, hours become more responsive to wage changes. This supports the claim that when young, because of the high human capital investment returns, labor supply is insensitive to the wage change. But when old, since the human capital effect is relatively insignificant, hours respond much more to wage changes. **?** noted that the human capital model implies this pattern.

7. SUMMARY AND CONCLUSIONS

In this article, we use the NLSY79 data to estimate the intertemporal elasticity of substitution in labor supply in a framework where people explicitly take into account human capital accumulation when they make labor supply decisions. We explicitly solve the continuous variable dynamic programming problem for optimal consumption and labor supply decisions and use the derived emax function in a ML routine. Using the estimated parameters, we conduct simulation experiments to generate age-wage, and age-labor hours profiles. We also use the simulated data to estimate the i.e.s. parameter using the conventional OLS and the IV methods.

The results indicate that the ML method based on the full solution of the continuous stochastic dynamic programming problem gives an estimated elasticity of intertemporal substitution parameter of 3.820, which is comparable to the elasticity results discussed in the macroliterature. In contrast, in the microliterature, MaCurdy and Altonji have obtained IV estimates using the PSID that range from roughly 0.37 to 0.88 . Using the NLSY79, and applying the same IV procedure as MaCurdy and Altonji, we obtain elasticity estimates of 0.260 using the raw data and 0.142 using data with outliers removed. We also find that if we simulate data from the structural model, with the substitution elasticity set to 3.82, and use conventional methods (IV) to estimate this parameter, we obtain estimates that are severely biased towards zero. Thus, the main reason for our much higher estimate of the intertemporal elasticity of substitution when we use the full solution procedure is clearly our explicit inclusion of human capital accumulation in the model.

The simulated age profiles of wages and the marginal rate of substitution between labor supply and consumption imply that in the early stage of the agents' careers, the effective wage, which we define as the marginal rate of substitution, is as much as 2.0 times higher than the real wage, implying that at younger ages, even if observed wages are low, the high effective wage resulting from high returns to human capital accumulation induces agents to have high labor supply. However, as agents acquire experience and become older, the ratio of the marginal rate of substitution to the wage falls. Through this mechanism, the labor supply model with human capital accumulation is able to reconcile a high elasticity of substitution with the fact that wages have a pronounced hump shape over the life cycle whereas the hump in hours is much more modest.

Finally, a word of caution is in order when one interprets the above results. Simulation results show that although the elasticity of intertemporal substitution was estimated to be around 3.8, the model does not imply that individuals change labor supply by a rate 3.8 times the rate of a wage change. On the contrary, the simulated hours response to a temporary wage increase of 2% ranged from 0.6% for young individuals to 4% for individuals near retirement. One reason is that when young, human capital accumulation is an important factor in determining labor supply, so temporary wage changes have little effect on labor supply. Second, even temporary wage shocks in the model here have some persistence, as increased labor supply leads to higher wages in the future through the human capital production function. Hence, there is an income effect.

APPENDIX

A.1. *Backward Solution of the Bellman Equations: Interpolation and Integration Steps.* At each asset and human capital grid point $\left(A_i, \tilde{K}_j\right)$ the value function is an integral over the taste shock and the wage shock. A straightforward way of performing these integrations is to use two-dimensional quadrature. That means, given the grid point $\left(A_i, \tilde{K}_j\right)$, calculate the quadrature points for the shocks,

$$\left(\varepsilon_{1,i_q}, \varepsilon_{2,j_q}\right) \quad i_q = 1, \ldots, n_q, \quad j_q = 1, \ldots, n_q$$

and then, solve for the optimal consumption and labor supply to get $V_{t,s}\left(A_t, \tilde{K}_t, \varepsilon_{1,i_q}, \varepsilon_{2,j_q}\right)$. Then, we can use quadrature to integrate over the value function to get the emax function. That is, form

$$V_{t,s}^{E}\left(A_i, \tilde{K}_j\right) = \iint V\left(A_i, \tilde{K}_j, \varepsilon_1, \varepsilon_2\right) d\varepsilon_1 d\varepsilon_2 \approx \sum_{i_q, j_q} V_{t,s}\left(A_i, \tilde{K}_j, \varepsilon_{i_q}, \varepsilon_{j_q}\right) w_{i_q} w_{j_q}$$

where $\left\{w_{i_q}\right\}$ are the weights for the quadrature integration.

But this approach is still computationally extremely demanding for several reasons, which are mainly due to the difficulties of applying two-dimensional quadrature integration. First, we still need to evaluate the value function over asset and human capital grid points, and wage and taste shock quadrature points. That means we need to conduct two-dimensional Newton search routines at $n_A \times n_K \times n_q \times n_q \times (T-20)$ points, where n_A, n_K are the number of asset and human capital grid points, and n_q is the number of quadrature points for taste shocks and human capital shocks. And it is the two-dimensional Newton search routine to find optimal consumption and labor supply which is by far the most computationally demanding part of the algorithm. Secondly, as we increase the dimension of the quadrature integration, even if we keep the quadrature points per dimension the same, we experience a decrease of the accuracy of the integration. Hence, if we wish to integrate over two dimensions and still have comparable accuracy to one-dimensional quadrature integration with n_q quadrature points, it is in general likely that we will need more than $n_q \times n_q$ quadrature points.

But these difficulties are minor compared to the problem of controlling for the range of K. Human capital $K = \tilde{K}\varepsilon_{1,i_q}$ can take on very small values if both $\tilde{K}$ and ε_{1,i_q} are small, and can take on very large values if both $\tilde{K}$ and ε_{1,i_q} are large. The Newton search routine at very low or high values of human capital is both disproportionately time consuming and inaccurate compared to Newton search at other points.

To avoid the quadrature integration with respect to the human capital shocks, we add another interpolation and approximation step that exploits the fact that there is a one-to-one mapping from human capital to wages. The basic logic is that the value function $V_{t,s}\left(A_t, \tilde{K}_t, \varepsilon_{1,t}, \varepsilon_{2,t}\right)$ is only a function of assets, A_t, human capital, K_t, and the taste shock $\varepsilon_{2,t}$. Once K_t is known, in order to calculate the value function, we do not need to know the values of the wage shock. So we go back to the original definition of the value function, in terms of $A_t, K_t = \tilde{K}_t\varepsilon_{1,t}$ and $\varepsilon_{2,t}$. The solution steps are as follows:

Step 1. Integrating the value function with respect to the taste shock. Assume that the age $t+1$ emax function

$$V_{t+1,s+1}^{E}\left(A_{t+1}, \tilde{K}_{t+1}\right) = E_t\left[V_{t+1,s+1}\left(A_{t+1}, \tilde{K}_{t+1}, \varepsilon_{1,t+1}, \varepsilon_{2,t+1}\right)\right]$$

is already calculated. We use Gauss-Hermite quadrature to integrate the expected value function over the taste shocks at asset and human capital grid points $\left(A_i, K_j\right)$. First, calculate the quadrature points and weights for the taste shock $\varepsilon_{2,t}$. Since $\varepsilon_{2,t}$ has a log normal distribution with parameters $\mu_2 = -\frac{1}{2}\sigma_2^2$ and $\sigma_2, \log\left(\varepsilon_{2,t+1}\right)$ is normally distributed with mean μ_2 and standard error σ_2. Let

$$x_{h,l}, \quad l = 1, \ldots, n_2$$

be the points for Gauss-Hermite quadrature. Then,

$$\varepsilon_{2,l}^q = \exp\left(\sqrt{2}\sigma_2 x_{h,l} + \mu_2\right), \quad l = 1, \ldots, n_2$$

are the Gauss-Hermite quadrature points for the above log normal distribution. Given $\left(A_i, K_j, \varepsilon_{2,l}^q\right)$, and the next period emax function, calculate the value function for each quadrature point of the taste shock $\varepsilon_{2,l}^q$ as follows:

$$V_{t,s}\left(A_i, K_j, \varepsilon_{2,l}^q\right) = \max_{\{C_t, h_t\}} \left\{u\left(C_t, t\right) - v\left(h_t, \varepsilon_{2,l}, t\right) + \beta V_{t+1,s+1}^E\left(A_{t+1}, \tilde{K}_{t+1}\right)\right\}$$

subject to the intertemporal budget constraint and human capital production function. Notice that this step requires a two-dimensional Newton search for optimal $\left(C_t, h_t\right)$ at only $n_A \times n_K \times n_q$ grid points, so the factor of n_q arising from the human capital shock is eliminated. Now, we can approximate the integration as follows using the quadrature procedure, with w^q being the weights for GaussHermite quadrature.

$$\begin{aligned}
& E_{\varepsilon_2} V_{t+1,s+1}\left(A_i, K_j \cdot \varepsilon_2\right) \\
& \quad = \int_{-\infty}^{\infty} V_{t+1,s+1}\left(A_i, K_j, \exp(z)\right) \frac{1}{\sqrt{2\pi\sigma_1}} \exp\left[-\frac{1}{2\sigma_1^2}\left(z - \mu_1\right)^2\right] dz \\
& \quad = \int_{-\infty}^{\infty} V_{t+1,s+1}\left(A_i, K_j, \exp\left(\sqrt{2}\sigma_1 x + \mu_1\right)\right) \frac{1}{\sqrt{\pi}} \exp\left[-x^2\right] dx \\
& \quad \approx \frac{\sum_{i=1}^{n_q} V_{t+1,s+1}\left(A_i, K_j, \varepsilon_{2,l}^q\right) w_l^q}{\sqrt{\pi}}
\end{aligned}$$

Step 2. Integrating the value function with respect to wage shock. Next, we integrate the value function with respect to the wage shocks to derive the emax function at the grid points $\left(A_i, \tilde{K}_j\right)$. From step 1 , for each grid point (A_i, K_j), we already have the value function integrated with respect to the taste shocks. Now, for each given asset grid value A_i, we fit Chebychev polynomials of log human capital to n_2 values of the integrated value function. That is, we derive

$$\hat{E}_{\varepsilon_2} V_{t,s}\left(A_i, K, \varepsilon_{2,t}\right) = \sum_{l=0}^{n_c} \pi_l T_l(\log(K))$$

where $T_l(\log(K))$ is the l th order Chebychev polynomial of log human capital. The coefficients π are derived by least squares with dependent variables being

$$E_{\varepsilon_2} V_{t,s}\left(A_i, K_j, \varepsilon_{2,t}\right), \quad j = 1, \ldots, n_2$$

Transform the Chebychev polynomials to the polynomials of $\log(K)$. Then,

$$\hat{E}_{\varepsilon_2} V_{t,s}(A_i, K, \varepsilon_{2,t}) = \sum_{i=0}^{n_c} \pi_i' \log(K)^i$$

Notice that for any realized human capital shock $\varepsilon_{1,t}$, the value function integrated over the taste shock is the sum of polynomials of the log wage shock and $\log(\tilde{K})$

$$\begin{aligned}
\hat{E}_{\varepsilon_2} V_{t,s}\left(A_i, \tilde{K}\varepsilon_{1,t}, \varepsilon_{2,t}\right) &= \sum_{i=1}^{n_c} \pi_i' \left[\log(\tilde{K}) + \log(\varepsilon_{1,t})\right]^i \\
&= \sum_{i=1}^{n_c} \pi_i' \sum_{j=0}^{i} \begin{pmatrix} n_c \\ j \end{pmatrix} [\log(\varepsilon_{1,t})]^j [\log(\tilde{K})]^{i-j}
\end{aligned}$$

Hence, we can integrate the value function with respect to the wage shock by integrating each wage shock polynomial separately as the following equation shows:

$$\begin{aligned}
E\left\{V_{t,s}\left(A_i, \tilde{K}\varepsilon_{1,t}, \varepsilon_{2,t}\right)\right\} &= E_{\varepsilon_1} \hat{E}_{\varepsilon_2} V_{t,s}\left(A_i, \tilde{K}\varepsilon_{1,t}, \varepsilon_{2,t}\right) \\
&= \sum_{i=1}^{n_c} \pi_i' \sum_{j=0}^{i} \begin{pmatrix} n_c \\ j \end{pmatrix} \left\{E_{\varepsilon_1} [\log(\varepsilon_{1,t})]^j\right\} [\log(\tilde{K})]^{i-j}
\end{aligned}$$

Here, since $\log(\varepsilon_{1,t})$ is normally distributed, integration of $[\log(\varepsilon_{1,t})]^j$ can be done analytically.

Then, we again approximate the above equation over $A_t, \tilde{K}_t$ using Chebychev polynomials to derive the emax function at age t, which we use for solving the age $t-1$ Bellman equation.

Notice that using the above algorithm, the problems of applying twodimensional quadrature are avoided. First, the Newton nonlinear search over consumption and labor needs only be applied to $n_A \times n_K \times n_q \times (T-20)$ points, hence reducing the computational time by a factor of n_q. Also, since only onedimensional quadrature integration is involved, there is no accuracy loss due to high-dimensional quadrature integration. Finally and most importantly, we only need to calculate the value function at the Chebychev grid point values for human capital and not at grid points for the human capital shock. Hence, we do not need to calculate value functions at extremely low or high human capital levels.

A.2. *Simulated Likelihood Calculation.* In this section, we describe the simulation steps that are required to construct the simulated likelihood function.

Step 1. Simulate $\{K_t^m, h_t^m, C_t^m, A_t^m\}_{t=t_0}^T$ starting from the initial period as follows:

1. Draw the true initial human capital $K_{t_0}^m$.

   First, draw the initial period measurement error ξ_0 and then, derive

$$K_{t_0}^m = \frac{K_{t_0}^D}{\xi_0}$$

2. Draw the true initial asset $A^m_{t_0}$.

   If the initial asset is observed, then draw the measurement error ξ_{3,t_0}, and derive

   $$A^m_{t_0} = A^D_{t_0} - \xi_{3,t_0}$$

   If the initial asset is not observable, then draw $A^m_{t_0}$ from $N(\bar{A}, V_{\bar{A}})$.

3. Draw the taste shock

   $$\ln(\varepsilon_2) \sim N(\frac{1}{2}\sigma_2^2, \sigma_2)$$

   and solve for the optimal consumption and labor supply. That is,

   $$\begin{aligned}\{C^m_{t_0}, h^m_{t_0}\} = \text{arg max}_{\{C_{t_0}, h_{t_0}\}}\{&u(C_{t_0}, t_0) - v(h_{t_0}, \varepsilon_{2,t_0}) + \beta E_{t_0} V_{t_0+1}[(1+r)A_{t_0} \\ &+ K^m_{t_0} h_{t_0} - C_{t_0}, \hat{K}_{t_0+1}, \varepsilon_{t_0+1}]\}\end{aligned}$$

   subject to

   $$\begin{aligned}A_{t_0+1} &= (1+r)A^m_{t_0} + K^m_{t_0} h_{t_0} - C_{t_0} \\ \hat{K}_{t_0+1} &= g(h_{t_0}, K^m_{t_0}, t_0)\end{aligned}$$

   Notice that we already have the polynomial approximation of the emax function

   $$V^E(A_{t_0+1}, \hat{K}_{t_0+1}) = E_{t_0} V_{t_0+1}(A_{t_0+1}, \hat{K}_{t_0+1}, \varepsilon_{1,t_0+1}, \varepsilon_{2,t_0+1})$$

   from the DP solution, which we will use in this case.

4. Draw the human capital shock ε_{1,t_0+1}, and derive the next period state variables.

   That is,

   $$\begin{aligned}A^m_{t_0+1} &= (1+r)A^m_{t_0} + K^m_{t_0} h^m_{t_0} - C^m_{t_0} \\ \hat{K}_{t_0+1} &= g(h^m_{t_0}, K^m_{t_0}, t_0) \\ K_{t_0+1} &= \hat{K}_{t_0} \varepsilon_{1,t_0+1}\end{aligned}$$

5. Now, repeat (3) and (4) until the end period T to derive the sequence of variables $\{K^m_t, h^m_t, C^m_t, A^m_t\}^T_{t=t_0}$.

Step 2. Given the simulated sequence of variables $\{K^m_t, h^m_t, C^m_t, A^m_t\}^T_{t=t_0}$, we then derive the measurement error. Then, we calculate the log likelihood increment for person i at the m^{th} simulation draw as follows.

Let us denote

$$\begin{aligned}\xi^m_0 &= \log K^D_{t_0} - \log K^m_{t_0} \\ \xi^m_{1,t} &= \log K^D_t - \log h^D_t - \log K^m_t - \log h^m_t \\ \xi^m_{2,t} &= h^D_t - h^m_t \\ \xi^m_{3,t} &= A^D_t - A^m_t\end{aligned}$$

If the initial wage at period to is available in the data, then we construct the initial wage measurement error $\tilde{\xi}_0^m$ and the log density of $\tilde{\xi}_0^m$ becomes part of the log likelihood increment of person i. If wage and hours data for person i at period $t > t_0$ is available, then we derive the labor income measurement error $\tilde{\xi}_1^m$, and the log density of $\tilde{\xi}_1^m$, becomes part of the log likelihood increment for person i at period t. On the other hand, if either wage or hours xor both for person i at period $t > t_0$ are not available, then the log likelihood increment for person i at period t does not contain any wage data K_t^D. Similarly, if hours data for person i at period t is available, then we derive the hours measurement error $\tilde{\xi}_2^m$, and the log density of $\tilde{\xi}_2^m$ becomes part of the log likelihood increment for person i at period t. On the other hand, if hours data are not available for person i at period t, then the log likelihood increment for person i at period t does not contain any hours information. Construction of the log likelihood increment for assets for person i at period t is done similarly.

Then, the log likelihood increment for person i is

$$\begin{aligned}
l_i^m = & \sum_{t=t_0+1}^{T} \left[\frac{(\tilde{\xi}_{1,t}^m + \frac{1}{2}\sigma_1)^2}{-2\sigma_1^2} - \log\sigma_1 - (\log K_t^D + \log h_t^d) \right] I(K_t^D, h_t^D \text{ observable }) \\
& + \sum_{t=t_0}^{T} \left[\frac{(\tilde{\xi}_{2,t}^m)^2}{-2\sigma_1^2} - \log\sigma_1 \right] I(h_t^D \text{ observable }) \\
& + \sum_{t=t_0+1}^{T} \left[\frac{(\tilde{\xi}_{3,t}^m)^2}{-2\sigma_3^2} - \log\sigma_3 \right] I(A_t^D \text{ observable }) \\
& + \left[\frac{(A_{t_0}^m - \bar{A})^2}{-2\sigma_{\bar{A}}^2} - \log\sigma_{\bar{A}} \right] I(A_t^D \text{ observable }) \\
& + \left[\frac{(\tilde{\xi}_0^m + \frac{1}{2}\sigma_0)^2}{-2\sigma_0^2} - \log\sigma_0 - \log K_{t_0}^D \right] I(K_t^D, h_t^D \text{ observable })
\end{aligned}$$

We set the starting time t_0 such that both $K_{t_0}^D$ and $h_{t_0}^D$ are observable.

Step 3. We repeat the simulation and likelihood increment calculation for $m = 1, \dots, M$ and derive the simulated log likelihood increment for individual i as follows:

$$l_i = \log\left[\sum_{m=1}^{M} \exp(l_i^m)\right]$$

The total log likelihood is

$$l = \sum_{i=1}^{N} l_i$$

A.3. *Data Generation.* We derived the wage, hours, and asset data from the NLSY as follows:

hours data: We use the variable "Number of hours worked in past calendar year" from 1979 to 1994. *wage data:* We first get total wage income data from the variable "Total Income from wages

and salary income past calendar year" from 1979 to 1994. And after adjusting for inflation using the GDP deflator, we divide the income variable by the hours variable to get the hourly wage rate.
asset data: We added up the following variables in the NLSY to construct total positive assets: "Total market value of vehicles including automobiles r/spouse own," "Total market value of farm/business/other property r/spouse own," "Market value of residential property r/spouse own," "Total market value of stocks/bonds/mutual funds," "Total amount of money assets like savings accounts of r/spouse," "Total market value of all other assets each worth more than $500."

We then added up the following variables to construct total negative assets: "Total amount of money r/spouse owe on vehicles including automobiles," "Total amount of debts on farm/business/other property r/spouse owe," "Amount of mortgages and back taxes r/spouse owe on residential property," "Total amount of other debts over $500 r/spouse owe."

The total amount of assets is calculated by subtracting the total amount of negative assets from the total amount of positive assets.

References