Lecture 1: Introduction to Household Decision Making and Human Capital and Life Cycle Models

DSE2021 University of Bonn

John Rust, Georgetown University

August 16, 2021

Thanks to David, Hans-Martin, and Philipp!

- They suggested the topic, raised the money, and took the risk to hold this summer school and conference
- Despite the Delta variant of Covid we have a fantastic line-up of lecturers and conference speakers who are leading experts on the topic of human capital and household decision making in life-cycle models
- Thanks to the hosts: University of Bonn and CRC/TR 224 grant from the EPoS Collaborative Research Center that made this possible
- Also thanks to the Econometric Society, and its sponsors Amazon, Two-Sigma Investments, and Google
- First DSE held in 2017 UCL hosted by Dennis Kristensen $\longrightarrow$ DSE2018 Copenhagen, Bertel Schjerning (machine learning)
- DSE2019 (IO and marketing) Günter Hitsch and Sanjog Misra, Chicago Booth
- DSE2020 (global warming) Felix Kübler and Simon Scheidegger, Zürich (unfortunately cancelled due to covid)

Road Map

1. What is “human capital”?
2. Why study household decision making?
3. Do we need life cycle models to study human capital?
4. Are we taking the life cycle model too literally?
5. Is it better to model decisions as discrete or continuous choices?
6. A brief and selective tour of life-cycle and human capital models
7. Are baby-boomers saving enough for their retirement?
8. What can neuroscience (and neuroeconomics) teach us about human capital?
9. Will all our human capital be worthless once we are replaced by robots with deep neural networks?

What is "human capital"?

- According to *Wikipedia* "Human capital is the stock of knowledge, skills, know-how and other acquired personal attributes (including good health and education) considered useful in the production process"
- Actually human capital is not just important for production, but also for personal happiness and well-being and overall social welfare and development
- It is an incredibly broad topic that involve that involves understanding how human capital evolves of the entire life-cycle, and is affected not only by one's one efforts and choices but also interactions in the family, in schooling and peer-groups as well as at work
- According to *Wikipedia* Adam Smith included in his definition of capital 'the acquired and useful abilities of all the inhabitants or members of the society'. The first use of the term 'human capital' may be by Irving Fisher. But the term only found widespread use in economics after its popularization by economists of the Chicago School, in particular Gary Becker, Jacob Mincer, and Theodore Schultz."

Challenges to studying human capital

- The study of human capital is challenging because skills/abilities/knowledge are encoded deep inside our brains and thus difficult to directly observe and measure. We have to rely on indirect indicators of human capital such as level of education, IQ/cognitive test scores, and wages.
- There is the age-old *nature versus nuture* question: how much of human capital (and success in life) is biologically inherited and how much is environmentally determined?
- *Many types of human capital and "skills"*: As Cunha, Heckmand and Schennach (2010 *Econometrica*) note "A large body of research documents the importance of cognitive skills in producing social and economic success. An emerging body of research estalishes the parallel importance of noncognitive skills, that is, personality, social, and emotional traits. Understanding the factors that affect the evolution of cognitive and noncognitive skills is important for understanding how to promote successful lives."

Human capital vs emotional and social capital

- According to *Wikipedia* "Today, most theories attempt to break down human capital into one or more components for analysis."
- *Emotional capital* is the set of resources (the personal and social emotional competencies) that is inherent to the person, useful for personal, professional and organizational development, and participates to social cohesion and has personal, economic and social returns (Gendron, 2004, 2008)."
- "*Social capital,* the sum of social bonds and relationships, has come to be recognized, along with many synonyms such as goodwill or brand value or social cohesion or social resilience and related concepts like celebrity or fame, as distinct from the talent that an individual (such as an athlete has uniquely) has developed that cannot be passed on to others regardless of effort, and those aspects that can be transferred or taught: instructional capital.

Challenges in studying human capital

- Increasingly *physical health* is also regarded as an important aspect or component of human capital
- *Wikipedia* notes that “Less commonly, some analyses conflate good instructions for health with health itself, or good knowledge management habits or systems with the instructions they compile and manage, or the ‘intellectual capital’ of teams – a reflection of their social and instructional capacities, with some assumptions about their individual uniqueness in the context in which they work. In general these analyses acknowledge that individual trained bodies, teachable ideas or skills, and social influence or persuasion power, are different.”
- How much of human capital is “pre-determined” by biology and early childhood environment and how much is “our responsibility” due to our own decisions and efforts/investments over our subsequent lifetimes?
- How much can *government policy* remediate inequality and promote human capital via education/training and family subsidies?

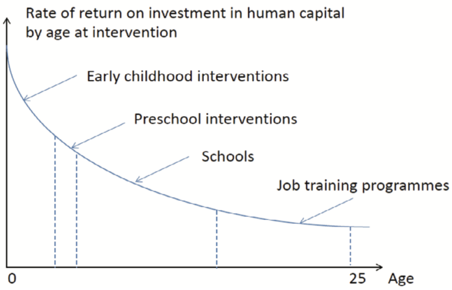
Why study household decision making?

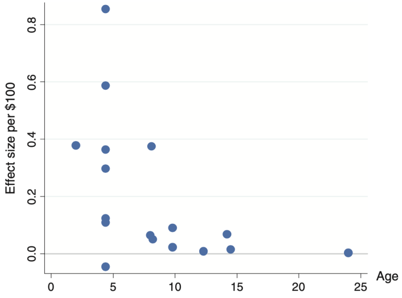
- Because parental decisions on how and where they raise their children has long-lasting effects on the childrens' human capital.
- As Attanasio, Nix and Meghir (2020) note "There is strong evidence showing that children's early experiences have long lasting effects, with implications for adult outcomes and even inter-generational transmission of human capital" "Our results show an important effect of early health on child cognitive development, which then becomes persistent. Parental investments affect cognitive development at all ages, but more so for younger children."
- Attanasio, Cunha and Jervis (2021) "Parental Beliefs and Investments in Human Capital" analyze "the importance of parental subjective beliefs in explaining the heterogeneity in parental choices of investments in the development of their children. Subjective beliefs about the production function of skills in early childhood development is crucial since parents may have biased beliefs about the returns to investments, which is crucial to pin down in designing policies aimed at remediating poor investments."

The "Heckman Curve"

- Cunha, Heckman and Schennach (2010) find that " Investments in the early years are important for the formation of adult cognitive skills. Furthermore, policy simulations from the model suggest that there is no trade-off between equity and efficiency. The optimal investment strategy to maximize aggregate schooling attainment or to minimize aggregate crime is to target the most disadvantaged at younger ages."
- Rosholm *et. al. Journal of Economic Surveys* (2021) 'Early intervention' has been a mantra in recent debates about human capital investment. Strong theoretical models motivate this focus by predicting that investment in children is most cost-effective when they are young."
- "This meta-analysis assembles cost-standardized effect estimates from 10 RCTs, including a total of 18 intervention arms and 30,578 participants (aged 1.5-24 years), conducted by the same research center in the Scandinavian welfare state of Denmark."

The “Heckman Curve” in Denmark





“Interventions targeted at younger children tend to produce larger effects, consistent with the Heckman curve. However, variation in the effect size within age groups is as large as it is across age groups. This indicates that both the quality and timing of investments matter and that ‘early interventions’ are not necessarily superior to later interventions.”

Do we need life cycle models to study human capital?

- In one sense, yes, because human capital evolves dynamically and stochastically over time, and actions/investments as well as subsequents events/outcomes affect how it evolves.
- But do we need to solve *structural dynamic programming models* involving decisions such as consumption-savings, labor/leisure, fertility, divorce, and the educational and other choices of parents and children and other parental investment decisions to properly understand the dynamics of human capital?
- **Answer:** Not necessarily. We can learn a lot from simpler, more agnostic models, but even those face daunting measurement and endogeneity issues.
- For example Cunha, Heckman and Schennach (CHS 2010) do not estimate a lifecycle model of household decision making but instead focus on estimating "the technology governing the formation of cognitive and non-cognitive skills in childhood"
- However the econometric challenge is that the inputs to this "human capital production function" are *endogenous* — the outcome of households' decision making processes.

Uncovering "causality" in human capital dynamics

- Their model has T periods of childhood followed by A periods of adult working life. Adult outcomes are produced by (latent) adult cognitive and non-cognitive skills $\theta_t = (\theta_{C,t}, \theta_{N,t})$ that were affected by parental investments $(I_{C,t}, I_{N,t})$ during childhood. Each agent is born (inherits) initial condition or "biological endowment" $\theta_1 = (\theta_{C,1}, \theta_{N,1})$ which may also reflect family environment, and parental skills $\theta_P = (\theta_{C,P}, \theta_{N,P})$. Let $\eta_t = (\eta_{C,t}, \eta_{N,t})$ denote random shocks to skills at time t. Then the *skill production function during childhood* is

$$\theta_{t+1} = f_s(\theta_t, I_t, \theta_P, \eta_t) \tag{1}$$

  where s denotes the *stage of child development.*
- CHS 2010 note that "Identifying and estimating technology (1) is challenging. Both inputs and outputs can only be proxied. Measurement error in general nonlinear specifications of technology (1) raises serious econometric challenges. Inputs may be endogenous and the unobservables in the input equations may be correlated with unobservables in the technology equations."

Adding health dynamics to human capital dynamics

- Attanasio, Meghir and Nix (2020) "builds on the nonlinear latent factor approach of Cunha, Heckman, and Schennach (2010) and uses the additional identification results from Agostinelli and Wiswall (2020) but is simpler and faster."
- "The inputs into the production functions include parental background, prior child cognition and health, and child investments, which are taken as endogenous. Estimation is based on a nonlinear factor model, based on multiple measurements for both inputs and child outcomes. Our results show an important effect of early health on child cognitive development, which then becomes persistent."
- "we find that health affects cognitive development, particularly at younger ages. The well-documented morbidity amongst poor populations, reflected in stunting, feeds into cognitive development generating permanent deficits in both health and cognition. These results imply that early health interventions are likely to be an important component in any policy that aims to improve human capital development."

Life-cycle models

- Life cycle models are even more ambitious in trying to infer the "production function" for human capital, but also closely related objects such as the *return to schooling* and *wage equation* and provide a model-based explanation/prediction of endogenous choices such as consumption, savings, labor supply, marriage/divorce decisios, fertility decisions, investment in children, decisions on application for social welfare benefits and retirement decisions and many many other possible decisions that we make over our life cycle.
- In principle, life cycle models can estimate and identify *subjective beliefs* about human capital accumulation and other causal relationship (e.g. subjective beliefs about the return to education, return to investing in children etc) that may differ from the "true" or "rational" *objective beliefs.*
- However there are two key constraints that limit our ability to model and learn about what real individuals and households do: 1) the curse of dimensionality, and 2) the identification problem.
- To do life-cycle modeling *be prepared to make and defend strong assumptions. You can't expect the the data to "speak for itself"*

A Canonical life-cycle model

- Consider first an *individual.* We formulate the life cycle model as a *single agent decision problem.* Later we will consider the complexity created by modeling decisions of a household, such as a married couple. Let s_t be a vector of *state variables* and d_t be a vector of *decision variables.* The individual's problem is

$$\max_{\{d_t\}_{t=0}^{\tilde{T}}} E\left\{\sum_{t=0}^{\tilde{T}} \beta^t u_t(s_t, d_t)\right\} + E\left\{\beta^{\tilde{T}} B(s_{\tilde{T}}, d_{\tilde{T}})\right\}. \tag{2}$$

  where $\tilde{T}$ is the random age of death and $b(\tilde{s}_{\tilde{T}})$ is a *bequest function.*
- Suppose there exists a T such that $\Pr\left\{\tilde{T} \leq T\right\} = 1$. Then we can solve the life cycle problem by backward induction from time T as follows.

Backward induction solution: the Bellman equation

- Define $V_T(s_T)$ as

$$V_T(s_T) = \max_{d_T \in D_T(s_T)} \left[u_T(s_T, d_T) + B(s_T, d_T) \right]. \tag{3}$$

where $D_T(s_T)$ is a set of feasible choices in state s_T. Define the *survival rate* $\rho_t(s_t) = \Pr\{\tilde{T} > t | s_t\}$. Then we can define $V_t(s_t)$ recursively as

$$V_t(s_t) = \max_{d_t} \left[u_t(s_t, d_t) + \beta \left(\rho_t(s_t) EV_{t+1}(s_t, d_t) + (1 - \rho_t(s_t)) B(s_t, d_t) \right) \right] \tag{4}$$

where

$$EV_{t+1}(s_t, d_t) = \int_{s_{t+1}} V_{t+1}(s_{t+1}) p_{t+1}(s_{t+1} | s_t, d_t) \tag{5}$$

where $p_{t+1}(s_{t+1}|s_t, d_t)$ is the individual's *subjective belief* about the probability of next period's state s_{t+1} given this period's state s_t and decision d_t.

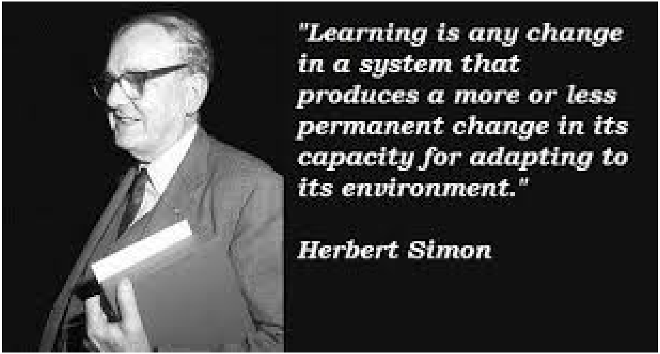
Taking the life-cycle model too literally

- There are some immediate problems with the life-cycle model outlined above as a literal model of individual behavior
- The *Allais Paradox* numerous laboratory experiments have rejected the *Independence Axiom* underlying expected utility, so there is strong evidence against the hypothesis that individuals' choices can be described by expected utility maximization.
- Ditto for discounted utility maximization. According to the 2002 review by Frederick, Loewenstein and O'Donohue, "The DU model, which continues to be widely used by economists, has little empirical support. Even its developers — Samuelson, who originally proposed the model, and Koopmans, who provided the first axiomatic derivation — had concerns about its descriptive realism, and it was never empirically validated as the appropriate model for intertemporal choice."
- Finally the rationality assumption is in doubt: Do individuals maximize their expected lifetime utility by backward induction or do more *ad hoc* adaptive learning processes more akin to *forward induction* provide a better description of how we make decisions over time?

The curse of dimensionality and "satisficing"

- Bellman realized a significant problem with dynamic programming of large scale, realistic DP problems with many state variables and decisions: *the curse of dimensionality.* Roughly speaking, the computer time required to solve a DP grows exponentially fast in the dimension (or number of) states and decisions. *We are only able to solve very simplified, and therefore unrealistic life cycle models.*
- Herbert Simon, in his 1978 Nobel prize speech, noted that "by the middle 1950s, a theory of bounded rationality had been proposed as an alternative to classical omniscient rationality [and] a significant number of empirical studies had been carried out that showed actual business decision making to conform reasonably well with the assumptions of bounded rationality but not with the assumptions of perfect rationality."

Herbert Simon



A defense of life cycle modeling

- The number of potential decisions we can choose from on any given day is astronomically large. Also our “state space” as encoded by our memories and senses of all the things that have happened to us and are currently going on is also astronomically large.
- Yet despite this vast set of possible states and decisions, and the fact that we are making decisions in continuous time, the average human being is able to function and function remarkably well. *And the human brain is able to do all of this running on approximately 20 watts of power.*
- Neuroscientists estimate that the calculations done by human brain, if they were translated and done on a digital computer, would be in excess of a *petaflop* i.e. 10^{15} floating point operations per second. Yet modern supercomputers that operate at this speed require hundreds of thousands of kilowatts.
- Is it possible that “mother nature” has discovered the key to breaking the curse of dimensionality via the *deep neural nets* in our mammalian brains?

Milton Friedman: the unbounded rationality hypothesis

- Friedman pushed the hypothesis that even if individuals and firms do not literally solve DP problems, they may act "as if" they had solved these problems, similar to the way expert professional pool players are making sophsticated use of the laws of physics (e.g. angle of incidence=angle of reflection, $MV = FT$, etc) having learned these laws through practice and trial and error even though they never took a physics course.
- Thus, it is not entirely crazy to use life-cycle models as a point of departure, even though we may have doubts about whether individuals are literally maximizing discounted expected utility by backward iduction.
- George Box: *All models are wrong, some are useful*
- For now, we will go with this aphorism since behavioral economics, psychology, computer science and neuroscience have not yet provided a computationally tractable alternative that can match the flexibility and generality of life cycle models that are solved using dynamic programming

Milton Friedman

Another issue: are decisions discrete or continuous?

- In life cycle models it is traditional to model *consumption* c_t as a *continuous choice* and individuals choose c_t subject to a budget constraint to maximize a continuously differentiable per-period utility function $u_t(c_t)$. Example

$$c_t = \underset{0 \le c \le W_t}{argmax} \left[u_t(c) + \beta \rho_t(W_t) EV_{t+1}(W_t - c) \right] \tag{6}$$

which implies that optimal consumption at time t satisfies the following *Euler equation*

$$0 = u_t'(c_t) + \beta \rho_t(W_t) EV_{t+1}'(W_t - c_t). \tag{7}$$

- But is this realistic? How often do we really make continuous consumption decisions as opposed to discrete decisions, e.g. ordering a particular item on the menu, or buying a carton of milk at the grocery store?
- Think of continuous choice as an approximation when we are only interested in the total amount spent on a large number of individual discrete purchases over a given time period such as a year.

Examples of life cycle models

- Ben-Porath (1967) *Journal of Political Economy* "The Production of Human Capital and the Life Cycle of Earnings"
- Let K_t be the stock of human capital at time t, and assume instantaneous earnings Y_t equals $Y_t = \alpha_0 K_t$.
- Let I_t be the amount invested in producing new human capital, so consumption is $C_t = Y_t - I_t$
- The production function for human capital is $Q_t = \beta_0 (s_t K_t)^{\beta_1} (D_t)_2^{\beta}$ where s_t is the fraction of current human capital devoted to producing more human capital (e.g. studying) and D_t are purchased inputs at price P (e.g. books, computers)
- Law of motion for human capital $\frac{d}{dt} K_t = Q_t - \delta K_t$ where $\delta \in (0, 1)$ is the depreciation rate of human capital
- $I_t = \alpha_0 s_t K_t + PD_t$ is the investment in human capital expressed in money terms
- The individual chooses an investment path to maximize the present value of earnings less costs of human capital investment

$$\max_{\{I_t\}} \int_0^T e^{-rt} \left[\alpha_0 K_t - I_t \right] dt \tag{8}$$

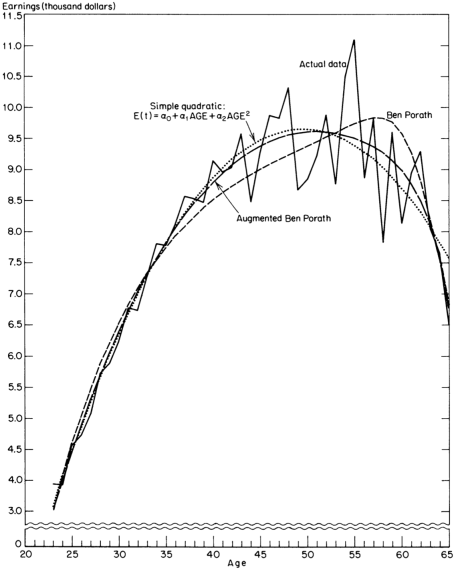
Heckman 1976

- Heckman 1976 *Journal of Political Economy* "A Life-Cycle Model of Earnings, Learning, and Consumption"
- Also a continuous time model, with $U(C_t, L_t H_t)$ where C_t is consumption, L_t is leisure and H_t is human capital at time t
- Human capital investment: $\frac{d}{dt}H_t = F(bI_t H_t, D_t) - \sigma H_t$. Earnings: $Y_t = RH_t$
- A_t are assets (wealth) at time t, $\frac{d}{dt}A_t = rA_t + Y_t[1 - L_t - I_t] - P_t D_t - C_t$
- The life cycle problem is

$$\max_{\{C_t, I_t, L_t\}} \int_0^T e^{-rt} U(C_t, L_t H_t) dt + B(A_T) \tag{9}$$

- Following the methodology of Ghez and Becker (1975), Heckman uses a discrete time version of his model to simulate the earnings profile of a typical individual predicted by his model.

Predicted vs actual wage profiles



Miller 1984

- Robert Miller 1984 *Journal of Political Economy* “Job Matching and Occupational Choice”
- Formulates a multi-armed bandit model of optimal Bayesian learning about the best occupation to maximize the expected discounted present values of wage earnings over their lifetime.
- Let $u_t(d)$ be the wage (including non-pecuniary benefits) of working in a job in occupation d at time t. Assume $u_t(d)$ is given by

$$u_t(d) = \xi_d + \sigma_d \epsilon_{d,t} \tag{10}$$

  where $\epsilon_{d,t} \sim N(0,1)$ is an *IID* idiosyncratic shock, σ_d is a known positive parameter and ξ_d is an unknown mean payoff to jobs in occupation d.
- The individual only observes realizations of the sum, $u_t(d)$ if they choose to work in a job in occupation d, but are Bayesian learners and their prior belief about ξ_d is that it is $\xi_d \sim N(\gamma_d, \delta_d^2)$.

- Let $h_t(d)$ be the total experience (think human capital) in occupation d at the start of period t, and d_s is the actual choice of occupation at time $s < t$, then

$$h_t(d) = \sum_{s=0}^{t-1} I\{d = d_s\}. \tag{11}$$

- By Bayes Rule, the posterior beliefs about ξ_d at time t are $N(\gamma_{d,t}, \delta^2_{d,t})$ where

$$\begin{aligned} \gamma_{d,t} &= \left[\gamma_d/\delta^2_d + \frac{1}{\sigma^2_d} \sum_{s=0}^{t-1} u_s(d_s) I\{d_s = d\} \right] \delta^2_{d,t} \\ \delta^2_{d,t} &= \left[\frac{1}{\delta^2_d} + \frac{h_t(d)}{\sigma^2_d} \right]^{-1} \end{aligned} \tag{12}$$

Miller 1984, continued

- The individual chooses a sequence of jobs to maximize expected discounted utility over his lifetime, taking the Bayesian learning into account
$$\max_{\{d_t\}} E\left\{\sum_{t=0}^{\infty} \beta^t u_t(d_t)\right\} \tag{13}$$
- This problem is known as a *multi-armed bandit problem* that was studied by Weitzman and Gittens and others. They discovered that the optimal search/learning strategy takes the form of a *reservation price rule* where $R(\gamma_{d,t}, \delta^2_{d,t})$ is the reservation price (expected payoff per unit of time) to working in occupation d. The formula for this reservation price is
$$R(\gamma_{d,t}, \delta^2_{d,t}) = \sup_{\tau \geq t}\left\{\frac{E_t\left\{\sum_{s=t}^{\tau} \beta^{s-t} u_s(d)\right\}}{E_t\left\{\sum_{s=t}^{\tau} \beta^{s-t}\right\}}\right\} \tag{14}$$
- The solution, τ is an *optimal stopping time* and roughly speaking, $R_d(\gamma_{d,t}, \delta^2_{d,t})$ is the expected payoff per unit of time of working in occupation d given the posterior beliefs $(\gamma_{d,t}, \delta^2_{d,t})$ about occupation d at time t.

Miller 1984, continued

- Thus, the optimal job search/occupation choice strategy is simple: at time $t = 0$ compute the reservation prices for each occupation, $\{R(\gamma_d, \delta_d^2)\}$ based on the *initial prior beliefs* about the returns (and variability of returns) to each occupation.
- Then choose the occupation d^* which has the highest reservation price $R(\gamma_{d^*}, \delta_{d^*}^2)$ and start working there.
- As the person works they learn about *that particular occupation* d^* and their reservation price changes over time as given above. If at any point their revised reservation price for occupation d^* falls below the reservation price for some other occupation d', then at that point it is optimal to switch from occupation d^* to occupation d' and the learning process resumes, but now in occupation d'.
- In a 1 occupation model, we can interpret ξ_d as the *match-specific effect of working at employer d* and the same reservation price strategy holds.

- This model implies a hazard rate $h_t(\alpha, \beta)$ which is the conditional probability of leaving the job (or occupation) after working on the job for t periods, where $\alpha = \sigma_d^2/\delta_d^2$ is the *information factor* (ratio of idiosyncratic noise to prior uncertainty about ξ_d).
- Let $p_t(\alpha, \beta)$ be the probability of working a given job t periods

$$p_t(\alpha, \beta) = h_t(\alpha, \beta) \prod_{s=1}^{t-1} [1 - h_s(\alpha, \beta)] \tag{15}$$

- If we observe a set of job spells $(t_1, \ldots, t_{N_i})$ for person i in a sample, the likelihood is

$$L = \prod_{i=1}^{N} \prod_{j=1}^{N_i} p_{t_j}(\alpha, \beta) \tag{16}$$

- "The empirical findings bolster the theoretical portion of the paper . . . since the value of specific experience varies across job types, optimizing behavior induces a stochastic career profile: jobs yielding returns that are subject to a particular form of uncertainty are experimented with first."

Keane and Wolpin 1997

- Keane and Wolpin (1997) *Journal of Political Economy* "Career Decisions of Young Men"
- They ignore the consumption/saving decision and assumption starting at age 16 young men make one of 5 possible decisions each period (1 year): 1) take blue collar job, 2) take white collar job, 3) join the military, 4) attend school, or 5) stay at home. They assume these choices are mutually exclusive, i.e. you cannot both attend school and stay at home.
- Let $u_a(d)$ be the utility obtained by an individual of age a from taking decision d. Keane and Wolpin assume that for the first three choices, $u_a(d)$ equals wage earnings computed as r_d, the "rental price" per unit of human capital, times the total human capital $h_a)d)$ given by

$$h_a(d) = \exp\{h_{16}(d) + \theta_1 g_a + \theta_2 x_a(d) - \theta_3 x_a(d)^2 + \epsilon_a(d)\} \quad (17)$$

where g_a is the total amount of schooling, $x_a(d)$ is the total amount of experience in sector d and $\epsilon_a(d)$ is a normally distributed random shock. Thus $u_a(d) = w_a(d) = r_d h_a(d)$.

Keane-Wolpin model, continued

- For going to school, $d = 4$, the utility is $u_a(4)$ given by

$$u_a(4) = h_{16}(4) - c_1 I\{g_a \geq 12\} - c_2 I\{g_a \geq 16\} + \epsilon_a(4) \qquad (18)$$

  and the utility of staying at home is

$$u_a(5) = h_{16}(5) + \epsilon_a(5) \qquad (19)$$

  where $\epsilon_a = (\epsilon_a(1), \epsilon_a(2), \epsilon_a(3), \epsilon_a(4), \epsilon_a(5))$ are *IID* $N(0, \Omega)$ shocks.

- Define the state of the individual at age a as $S_a = (e_{16}, g_a, x_a, \epsilon_a)$ where $x_a = (x_a(1), x_a(2), x_a(3))$ is total experience in blue collar, white collar and the military. We have

$$V_a(S_a) = \max_{d_a} E\left\{ \sum_{t=a}^{A} \delta^{t-a} u_a(d) \right\} \qquad (20)$$

  Let $E_a = (e_{16}, g_a, x_a)$ be the observed *experience* state, so $S_a = (E_a, \epsilon_a)$ where ϵ_a is the *unobserved state.*

Keane-Wolpin model, continued

- *Conditional choice probability* Let $P(d_a|E_a)$ be the probability a person with observed experience E_a will choose alternative d_a at age a. This can be computed by (numerical) integration over the unobserved state variables ϵ_a

$$P(d_a|E_a) = \int_{\epsilon_a} I\{\delta_a(E_a,\epsilon_a) = d_a) = d_a\}\phi(\epsilon_a)d\epsilon_a \qquad (21)$$

- Let θ be the unknown parameters of the model. Using the conditional choice probability we can form a likelihood function using *panel data* $\{(d_a, E_a)|a = \underline{a}_i, \ldots, \overline{a}_i\}$ on the choices and experience histories of a sample of N young men.

$$L(\theta) = \prod_{i=1}^{N} \prod_{a=\underline{a}_i}^{\overline{a}_i} P(d_a|E_a,\theta) \qquad (22)$$

- They estimated the model using a sample of 1373 young men (followed for up to 11 years) using the NLSY survey starting in 1979.

Keane-Wolpin model, continued

- "We find that an augmented human capital investment model does a good job of fitting the data on the educational and occupational choices of this cohort. The model, however, is a considerable extension beyond a 'bare-bones' human capital investment model."
- "Of particular importance for fitting the data was the inclusion of skill depreciation during periods of nonwork, of mobility or job-finding costs, of school reentry costs, and of nonpecuniary components of occupational payoffs."
- "A more parsimonious model, which allowed only for occupation-specific human capital accumulation (occupation-specific work experience), general human capital accumulation (schooling), and unobserved endowment heterogeneity, but did not contain these additional elements, could not explain either the degree of persistence in occupational choices or the rapid decline in schooling with age."

Rust and Phelan 1997

- Rust and Phelan 1997 *Econometrica* "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets"
- Models retirement behavior and the impact of health and Social Security on retirement at the end of the life cycle
- Ignores the consumption/saving decision and assumes $c_t = y_t$ with probability 1. The rationale for doing this is twofold:
- (a) consumption is very difficult to measure in the RHS dataset and attempts to impute it from the budget constraint $c_t = y_t + w_t - w_{t+1}$ (where w_t, is the individual's net worth at time t) yielded implausibly erratic consumption paths and a disturbingly high incidence of negative measured consumption (see Rust (1990) for details), and
- (b) our predominantly blue-collar subsample of the RHS has no significant tangible wealth beyond housing equity.
- Total net worth amounts to less than five years of income, and home equity consists of over 50% this net worth.

A generic life cycle model with extreme value shocks

- Let s_t be a vector of state variables at time t and d_t a vector of decision variables. Assume $s_t = (x_t, \epsilon_t)$ where we (the econometrician) can observe x_t but only the agent (individual, household, etc) observes ϵ_t.
- Let the utility function $u_t(s, d, \theta_u)$ take the additively separable form $u_t(x, d) + \epsilon_t(d)$ where $\epsilon_t(d)$ is the component of utility at time t associated with the decision d and states that the econometrician does not observe. We call this *Assumption AS* (Additive Separability)
- The value function is

$$V_t(s) = \max_{\delta} E_\delta \left\{ \sum_{j=t}^{T} \beta^{j-t} u_j(s_j, d_j, \theta_u) | s_t = s \right\} \tag{23}$$

where $\delta = (\delta_1, \ldots, \delta_T)$ is a *policy* (or sequence of *decision rules*) where $d_t = \delta_t(s_t)$ specifies the action the individual takes in state s_t. In addition we require *feasibility* so $\delta_t(x, \epsilon) \in D_t(x)$ for all t and $s = (x, \epsilon)$ where $D_t(x)$ is a *finite state-dependent choice set* (budget set).

A generic life cycle model with extreme value shocks

- The second key assumption we make is *Assumption CI* (Conditional Independence). $\{x_t, \epsilon_t\}$ is a controlled Markov process with transition densities

$$\begin{aligned} p_{t+1}(s_{t+1}|s_t, d_t) &= p_{t+1}(x_{t+1}, \epsilon_{t+1}|x_t, \epsilon_t, d_t) \\ &= q(\epsilon_{t+1}|x_{t+1})p_{t+1}(x_{t+1}|x_t, d_t) \end{aligned} \tag{24}$$

- This implies that ϵ_{t+1} does not directly depend on ϵ_t: at most the dependence is indirect if ϵ_t affect d_t which then affects s_{t+1} which is a conditioning variable in $q(\epsilon_{t+1}|x_{t+1})$.
- The final assumption is *Assumption EV* (Extreme Value): the CDF $q(\epsilon_t|x_t)$ is given by

$$q(\epsilon|x) = \prod_{d \in D(x)} \exp\{-\exp\{-(\epsilon(d) - \mu(d))/\sigma\}\} \tag{25}$$

where $\sigma > 0$ is a scale parameter and $\mu(d)$, $d \in D(x)$ are location parameters.

What does AS, CI and EV buy us?

- Consider Bellman's equation for solving the life cycle problem. By Assumption AS we have

$$V_t(s) = V_t(x,\epsilon) = \max_{d \in D(x)} \left[u(x,d,\theta_u) + \epsilon(d) + \beta EV_{t+1}(x,\epsilon,d)\right] \tag{26}$$

  and by Assumption CI we have

$$\begin{aligned} EV_{t+1}(x,\epsilon) &= \int_{x'}\int_{\epsilon'} V_{t+1}(x',\epsilon')p_{t+1}(x',\epsilon'|x,\epsilon,d) \\ &= \int_{x'}\int_{\epsilon'} V_{t+1}(x',\epsilon')q(\epsilon'|x')p_{t+1}(x'|x,d). \end{aligned} \tag{27}$$

- Notice that $EV_{t+1}(x,\epsilon,d)$ thus actually only depends on (x,d), i.e. $EV_{t+1}(x,d)$. Define $v_t(x,d)$ by

$$v_t(x,d) = u_t(x,d) + \beta EV_{t+1}(x,d) \tag{28}$$

What do Assumptions AS and CI buy us?

- So we conclude that Assumptions AS and CI imply the following representation holds

$$V_t(x, \epsilon) = \max_{d \in D(x)} [v_t(x, d) + \epsilon(d)] \tag{29}$$

- If follows that even though the choice problem is dynamic, the choice probabilities implied by the dynamic discrete choice model are isomorphic to choice probabilities from a static discrete choice model with utility function $v_t(x, d)$

$$P_t(d|x) = \int_{\epsilon} I\{d = \delta_t(x, \epsilon)\}\, q(\epsilon|x) \tag{30}$$

where $\delta_t(x, \epsilon)$ is the optimal decision rule,
$\delta_t(x, \epsilon) = argmax_{d \in D(x)} [v_t(x, d) + \epsilon(d)]$.

What does Assumption EV buy us?

- Analytic expressions for the "EMAX" and choice probabilities.

$$\begin{aligned} EV_{t+1}(x,d) &= \int_{x'}\int_{\epsilon'} V_{t+1}(x',\epsilon')q(\epsilon'|x')p_{t+1}(x'|x,d) \qquad (31) \\ &= \int_{x'} \sigma \log\left(\sum_{d'\in D(x')} \exp\{v_{t+1}(x',d')/\sigma\}\right) p_{t+1}(x'|x,d) \end{aligned}$$

  where the inner expectation with respect to ϵ' has the closed-form "log-sum" expression.
- Further, the EV assumption results in closed-form expressions for the conditional choice probabilities (CCPs)

$$\begin{aligned} P_t(d|x) &= \int_{\epsilon} I\{d = \delta_t(x,\epsilon)\}\, q(\epsilon|x) \\ &= \frac{\exp\{v_t(x,d)/\sigma\}}{\sum_{d'\in D(x)} \exp\{v_t(x,d')/\sigma\}} \qquad (32) \end{aligned}$$

  Note that under the EV Assumption the CCPs have the multinomial logit form (MNL) just as in static discrete choice models that adopt the AS and EV assumptions.

Recap

- The original Bellman equation involved backward induction to compute the sequence $\{V_t(x,\epsilon)\}$ of value functions

$$V_t(x,\epsilon) = \max_{d\in D(x)} \left[u_t(x,d) + \beta \int_{x'} \int_{\epsilon'} V_{t+1}(x',\epsilon') p_{t+1}(x',\epsilon'|x,\epsilon,d) \right] \tag{33}$$

- Assumptions AS, CI and EV imply that we can instead solve the DP problem for the sequence $\{v_t(x,d)\}$ via the recursion

$$v_t(x,d) = u_t(x,d) + \tag{34}$$
$$\beta \int_{x'} \sigma \log \left(\sum_{d'\in D(x')} \exp\{v_{t+1}(x',d')/\sigma\} \right) p_{t+1}(x'|x,d)$$

- This second recursion is much easier and faster since we do not have to numerically integrate with respect to $q(\epsilon'|x')$ since the log-sum formula has already done that for us. Note: these integrals have to be done many times: for each x and for each $t \in \{1,\ldots,T\}$.
- Also the closed-form MNL expression for CCPs saves us additional multivariate integrations!

Maximum likelihood estimation

- Let $\theta = (\beta, \theta_u, \theta_p)$ be the vector of parameters to be estimated, where β is the discount factor, θ_u are the parameters determining the utility functions, $\{u_t(x, d, \theta_u)\}$ and θ_p are the parameters determining beliefs, $\{p_t(x'|x, d, \theta_p)\}$.
- Suppose we have *unbalanced panel data* $\{(x_{it}, d_{it}|t \in (\underline{T}_i), \ldots, \overline{T}_i,\ i = 1, \ldots, N\}$ following the observed choices and states of N different individuals. Assume the individuals' realized states and decisions are independent of each other. But the sequence of decisions is clearly serially correlated.
- Let $L_f(\theta)$ be the full likelihood function given by

$$L_f(\theta) = \prod_{i=1}^{N} \prod_{t=\underline{T}_i}^{\overline{T}_i} P_t(d_{it}|x_{it}, \theta) p_t(x_{it}|x_{it-1}, d_{it-1}, \theta_p) \qquad (35)$$

where (x_{i0}, d_{i0}) are taken as fixed *initial conditions* for each individual $i \in \{1, \ldots, N\}$.

Two (or Three) step estimation

- If we assume *rational expectations* we can estimate θ_p in a first stage using the *partial likelihood function* $L_p(\theta_p)$
- **Step 1** estimate θ_p using $L_p(\theta_p)$. Note that estimation of θ_p does not require any nested solution of the life cycle problem.

$$L_p(\theta_p) = \prod_{i=1}^{N} \prod_{t=\underline{T}_i}^{\overline{T}_i} p_t(x_{it}|x_{it-1}, d_{it-1}, \theta_p) \tag{36}$$

- **Step 2** Estimate the remaining parameters using $L_u(\beta, \theta_u, \hat{\theta}_p)$ using the Step 1 estimates $\hat{\theta}_p$ as the true values.

$$L_u(\beta, \theta_u, \theta_p) = \prod_{i=1}^{N} \prod_{t=\underline{T}_i}^{\overline{T}_i} P_t(d_{it}|x_{it}, \beta, \theta_u, \hat{\theta}_p) \tag{37}$$

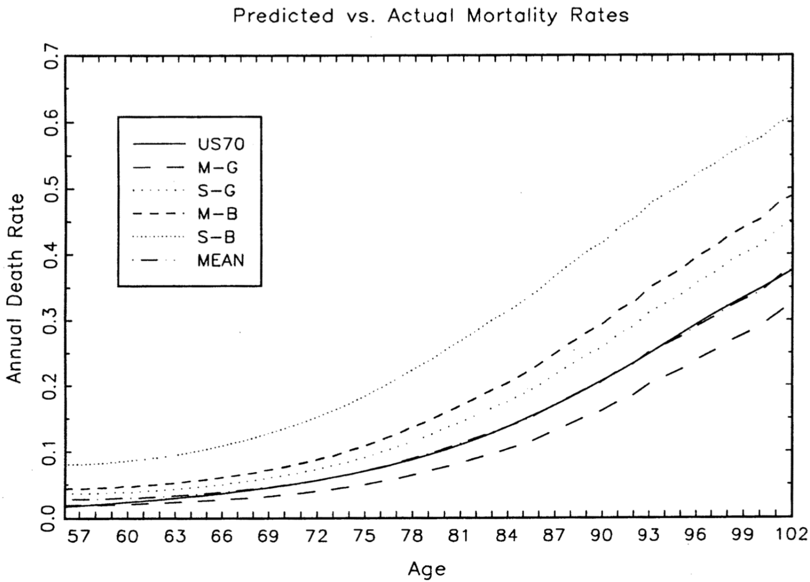
- **Step 3** (optional) Do 1 Newton step using the full likelihood $L_f(\theta)$ starting from the Step 2 estimates $(\hat{\beta}, \hat{\theta}_u, \hat{\theta}_p)$ to get fully efficient estimates and correct standard errors.

First stage health transition probabilities

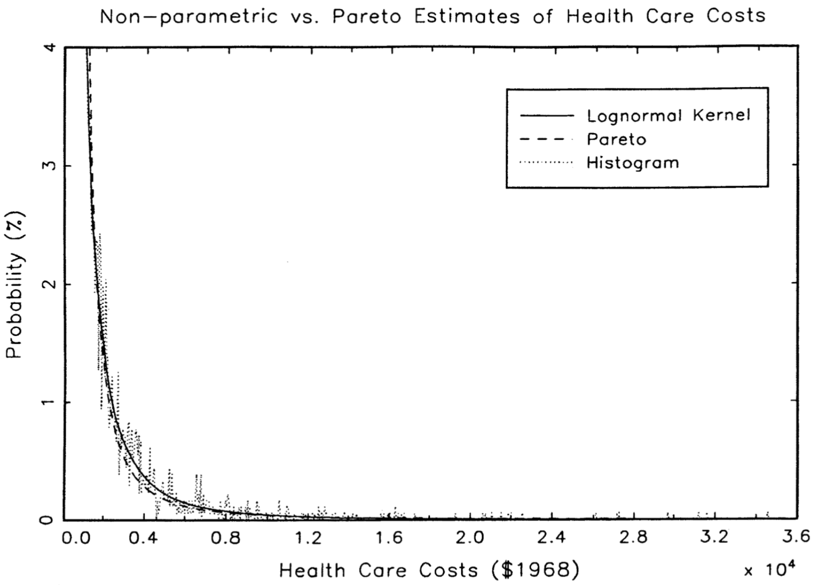
EXAMPLES OF HEALTH STATUS TRANSITION PROBABILITIES

Current Health Status	Current wage and Future Health Status			
	Low Wage		High Wage	
	G	B	G	B
Age 56:				
G	80.2	19.8	91.5	8.5
B	34.3	65.7	58.1	41.9
Age 64:				
G	71.6	28.4	87.0	13.0
B	24.5	75.5	46.4	53.6
Age 80:				
G	49.4	50.6	72.2	27.8
B	11.2	88.8	25.1	74.9

First stage mortality rates



First stage Pareto health care costs



Model fit, not yet receiving social security

PREDICTED VS. ACTUAL CHOICE PROBABILITIES FOR MEN AGED 62–63 NOT ALREADY RECEIVING SOCIAL SECURITY

$e = FT$, $N = 1221$ $\chi^2 = 20.57$	FT, DB	PT, DB	NE, DB	FT, AB	PT, AB	NE, AB
NP	30.96	1.06	0.25	49.88	14.09	3.77
DP	30.15	1.89	0.63	49.70	11.91	5.73

$e = FT$, $hi = eph$, $N = 1065$ $\chi^2 = 29.22$	FT, DB	PT, DB	NE, DB	FT, AB	PT, AB	NE, AB
NP	31.46	0.94	0.09	50.99	14.08	2.44
DP	31.12	1.68	0.41	50.79	11.00	5.00

$e = FT$, $hi \in \{gph, mca, nhi\}$, $N = 156$ $\chi^2 = 4.64$	FT, DB	PT, DB	NE, DB	FT, AB	PT, AB	NE, AB
NP	27.56	1.92	1.28	42.31	14.10	12.82
DP	23.46	3.31	2.16	42.21	18.14	10.71

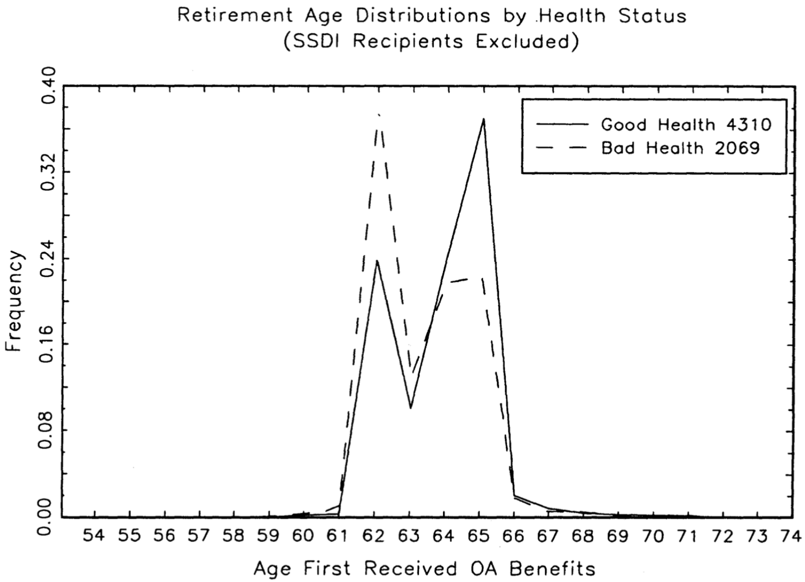
$e = FT$, $h = B$, $N = 270$ $\chi^2 = 11.01$	FT, DB	PT, DB	NE, DB	FT, AB	PT, AB	NE, AB
NP	26.67	1.48	0.74	48.52	17.41	5.19
DP	28.14	2.03	1.04	46.84	12.59	9.36

Model fit, age 65 or older

SUMMARY OF THE INCENTIVE EFFECTS OF SOCIAL SECURITY AND MEDICARE ON MALE LABOR SUPPLY

$e = FT$, $t > 65$, $ss \in \{ER, NR\}$, $N = 1077$; $\chi^2 = 0.70$	FT	PT	NE
NP	52.55	22.19	25.26
DP	52.61	23.05	24.34
$e = FT$, $t \geq 65$, $ss = NE$, $N = 289$; $\chi^2 = .10$	FT	PT	NE
NP	70.69	16.26	13.15
DP	69.41	17.09	13.50
$e \in \{FT, PT\}$, $t \geq 65$, $ss \in \{ER, NR\}$, $N = 1767$; $\chi^2 = .47$	FT	PT	NE
NP	36.67	34.47	28.86
DP	37.43	34.23	28.34
$e \in \{FT, PT\}$, $t \geq 65$, $ss = NE$, $N = 312$; $\chi^2 = .08$	FT	PT	NE
NP	66.67	18.91	14.42
DP	65.92	19.50	14.58
$e \in \{FT, PT, NE\}$, $t \geq 65$, $ss \in \{ER, NR\}$, $N = 3445$; $\chi^2 = .74$	FT	PT	NE
NP	19.30	20.00	60.70
DP	19.81	20.16	60.03
$e \in \{FT, FT\}$, $t \geq 65$, $ss = NE$, $N = 334$; $\chi^2 = .18$	FT	PT	NE
NP	62.28	18.87	18.86
DP	61.70	18.33	19.77

Distributions of retirement ages by health status



Conclusions on the effect of Social Security

- How do Social Security and Medicare affect retirement behavior in a world of incomplete markets?
- Our results suggest that Social Security creates significant disincentives to labor force participation, and is largely responsible for the peaks in retirements at ages 62 and 65, the ages of eligibility for early and normal Social Security retirement benefits
- The DP model allows us to conduct the conceptual experimentof comparing the behavior of two otherwise identical individuals, one of whom is entitled to Social Security benefits. We found that an individual who is employed and entitled to Social Security benefits is significantly less likely to continue working than his counterpart who is not entitled
- Conversely an unemployed individual who is entitled to Social Security has very little chance of returning to work in comparison to an identical individual who is not entitled.
- The most radical experiment — eliminating Social Security entirely — ends up completely eliminating the peaks in retirements at ages 62 and 65.

French (2005)

- Eric French (2005) *Review of Economic Studies* “The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour”
- First paper to structurally estimate a retirement model that both models labor supply, retirement (Social Security claiming decisions) and the consumption/savings decision.
- “Previous structural analyses of labour supply and retirement behavior have made diametrically opposed assumptions about a household’s ability to borrow and save. At one extreme, Burtless (1986) and Gustman and Steinmeier (1986) assume that households perfectly smooth consumption by borrowing and lending without limit. At the opposite extreme Stock and Wise (1990) and Rust and Phelan (1997) assume that households cannot borrow or save, thus allowing no intertemporal consumption smoothing. Clearly, neither of the assumptions is correct.”

French's Model

- Models the life cycle consumption/saving, participation/hours worked and pension/Social Security claiming decisions from age 30 to 95, estimating the model using PSID data.
- Accounts for health (good and bad health), assets, wages, AIME, and spousal income/benefits.
- Uses non-additively separable CRRA/Cobb-Douglas utility function

$$U(C,H,M) = \frac{1}{1-\nu}\left(C_t^{\gamma}[L - H_t - \theta I\{H_t > 0\} - \phi I\{M = \text{bad}\}]^{1-\gamma}\right)^{1-\nu} \tag{38}$$

and bequest function

$$b(A_t) = \theta_B \frac{(A_t + K)^{(1-\nu)\gamma}}{1-\nu}. \tag{39}$$

and wages with an autoregressive error term

$$W_t = \alpha \log(H_t) + W(M_t) + u_t, \quad u_t = \rho u_{t-1} + \eta_t,\ \eta_t \sim N(0, \zeta_\eta^2) \tag{40}$$

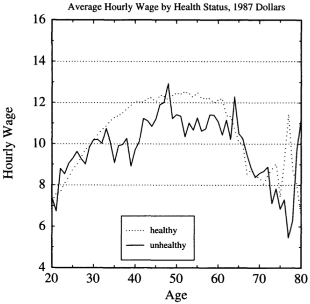
Solution of French's Model

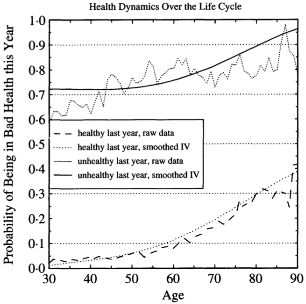
- "Because the fixed cost of work and the benefit application decision mean that the value function need not be globally concave, I cannot use relatively fast hill-climbing algorithms. Therefore, I also discretize the consumption and labour supply decisions and use a grid search technique to find the optimal consumption and hours rules. The grids seem to produce reasonable approximations."
- "the state variables are discretized into a finite number of points on a grid and the value function is evaluated at those points. I use linear interpolation with the grid and extrapolation outside of the grid to evaluate the value function computed. I integrate the value function with respect to the innovation in the wage using quadrature."
- "The model solution procedure allows for heterogeneity via the state variables. However, the requirement of computational simplicity does not allow for heterogeneity in preferences or in the data generating process for the state variables."

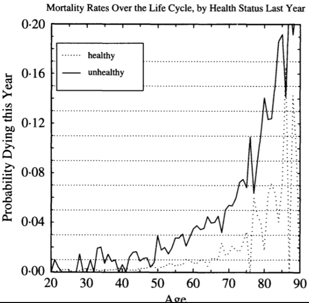
Solution of French's Model

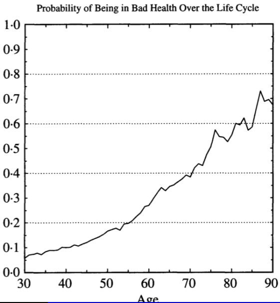
- "The MSM estimation strategy matches mean assets, hours of work, participation, and also median assets in the PSID to the corresponding moments of the same variables in a simulated sample. The 'matching' of moments is done using standard GMM techniques. Because of prob measurement error, I do not match high order moments. Using means, however, averages out measurement error, as shown below."
- "I pick an arbitrary vector of preference parameters and compute the decision rules given those parameters and the numerical methods described in Section 2.4. Then I use the decision rules and the health and wage shocks to simulate hypothetical life cycle profiles for the decision variables. Finally, the simulated data and the true data are aggregated by age (and in the case of participation, by health status). Then, the difference between the simulated and true profiles is computed and the differences are weighted up to form a distance measure. Then a new vector of preference parameters is picked and the whole process is repeated until we find $\hat{\theta}$, the MSM esimates that minimize the distance between the data moments and simulated moments."

Life cycle profiles of labor supply and assets

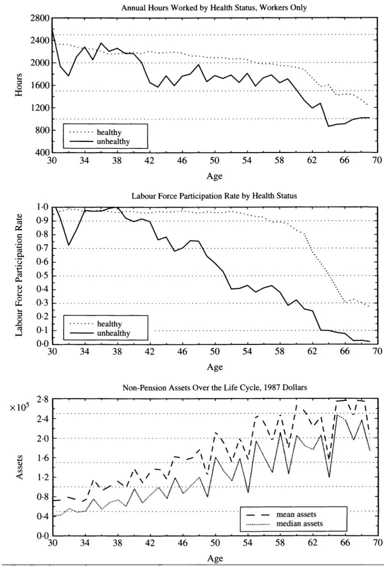








Life cycle profiles of labor supply and assets



Structural parameter estimates from MSM

Preference parameter estimates

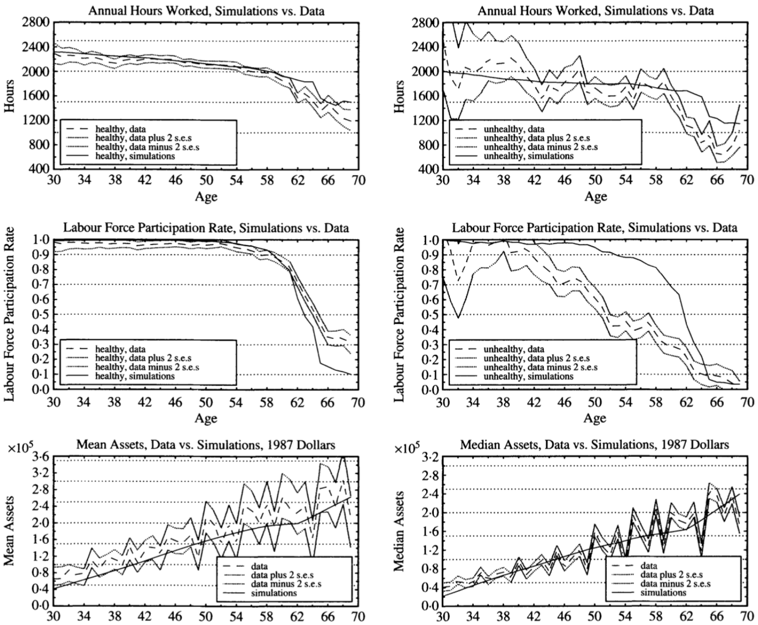
Parameter and definition	Specification (1)	(2)	(3)	(4)
γ Consumption weight	0·578 (0·003)	0·602 (0·003)	0·533 (0·003)	0· 615 (0·004)
ν Coefficient of relative risk aversion, utility	3·34 (0·07)	3·78 (0·07)	3·19 (0·05)	7·69 (0·15)
β Time discount factor	0·992 (0·002)	0·985 (0·002)	0·981 (0·001)	1·04 (0·004)
L Leisure endowment	4466 (30)	4889 (32)	3900 (24)	3399 (28)
ϕ Hours of leisure lost, bad health	318 (9)	191 (7)	196 (8)	202 (6)
θ_P Fixed cost of work, in hours	1313 (14)	1292 (15)	335 (7)	240 (6)
θ_B Bequest weight	1·69 (0·05)	2·58 (0·07)	1·70 (0·04)	0·037 (0·001)
χ^2 Statistic: (233 degrees of freedom)	856	880	830	1036
$\epsilon_{h,W}(40)$ Labour supply elasticity, age 40	0·37	0·37	0·35	0·19
$\epsilon_{h,W}(60)$ Labour supply elasticity, age 60	1·24	1·33	1·10	1·04
Reservation hours level, age 62	885	916	1072	1051
Coefficient of relative risk aversion	2·35	2·68	2·17	5·11

Standard errors in parentheses

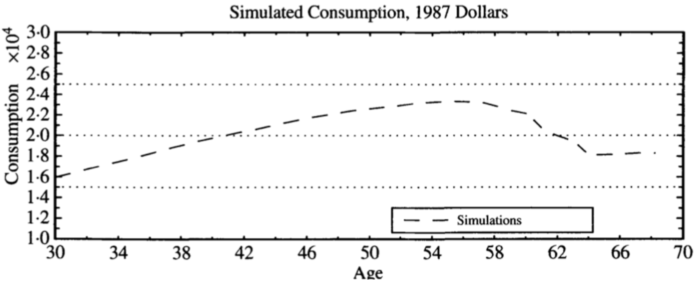
Specifications described below:

(1) Does not account for selection or tied wage-hours offers
(2) Accounts for selection but not tied wage-hours offers
(3) Accounts for tied wage-hours offers but not selection
(4) Accounts for selection and tied wage-hours offers

Comparison of Model Predictions to Data



Mean Simulated Consumption Path



- Notice the drop in consumption around retirement (ages 58 to 64). Is this drop inconsistent with consumption smoothing? "a number of empirical studies (e.g. Banks et al. 1998, Bernheim et al. 2001, Haider and Stephens 2007 and Schwerdt 2005) found a sharp decline in consumption during the first years of retirement, a phenomenon referred to as the *retirement-consumption puzzle.* It is puzzling to economists why households do not plan properly and save enough for an expected fall in income." Olafsson and Pagel (2018) "The retirement-consumption puzzle: New evidence from personal finances"

Policy experiments

Policy experiments, $\theta_B = 0$, $\beta = 0{\cdot}95$, $\frac{-1}{\gamma(1-\nu)-1} = 0{\cdot}48$

	Years worked	Hours worked per year	PDV of labour income ($)	PDV of consumption ($)	Assets at age 62 ($)
With borrowing constraints					
Current policies	36·77	2003	1788	1840	10
Reduce benefits 20%	37·42	2000	1805	1828	13
Reduce benefits 20%, reduce taxes	37·66	2011	1819	1840	14
Shift early retirement age to 63	36·77	2003	1788	1840	11
Eliminate earnings test, age 65+	37·91	2005	1811	1858	8

PDV is present discounted value.
Consumption, labour income, and assets are measured in thousands.

- Notice that raising the Social Security retirement age from 62 to 63 has almost no effect. "Recall that benefit recomputation formulae almost fully replace benefits lost through the earnings test at age 62. Therefore, if borrowing constraints do not bind, there should be little if any work disincentive imposed by Social Security at age 62 and thus there should be little if any effect of shifting the Social Security early retirement age to 63. Recall that borrowing constraints bind for very few individuals at age 62. As a result, any effect of this policy would be minor"

Comments on the policy experiments

- It may be important to account for different life cycle dynamics of individuals with different levels of human capital, dynamics that cannot be adequately captured by a "homogeneous" model, even though there is heterogeneity induced by the realizations of state variables over the life cycle.
- Not clear that all individuals are behaving like rational life-cycle savers. According to longitudinal study by the Economic Policy Institute "Half of American families in the 56-to-61 age bracket had less than $21,000 in retirement savings in 2016" and "Forty percent of Americans over the age of 60 who are no longer working full-time rely solely on Social Security for their income and the median annual benefit is about $17,000."
- According to an article in the *Washington Post* "Every day 10,000 Americans turn age 65. And every year, fewer and fewer of them have traditional employer-sponsored pensions to support them. The system that was supposed to provide for them is shot through with holes."

Are people saving adequately for retirement?

- Contrast this with Scholz, Sheshadri and Khitratrakun 2006 *Journal of Political Economy* “Are Americans Saving Optimally for Retirement?”
- “We solve each household’s optimal saving decisions using a life cycle model that incorporates uncertain lifetimes, uninsurable earnings and medical expenses, progressive taxation, government transfers, and pension and social security benefits. With optimal decision rules, we compare, household by household, wealth predictions from the life cycle model using a nationally representative sample.”
- “We find, making use of household-specific earnings histories, that the model accounts for more than 80 percent of the 1992 cross-sectional variation in wealth. Fewer than 20 percent of households have less wealth than their optimal targets, and the wealth deficit of those who are under-saving is generally small.”
- However studied the HRS 1931-41 birth cohort. Many of these individuals could have been *scarred* by the Great Depression. The baby boomers suffered a different type of scarring — too much *sex, drugs, and rock and roll*