

Retiring Cold Turkey

Xiaodong Fan*
CEPAR, UNSW

October 29, 2012

Abstract

This paper documents “sharp retirement”—retirement accompanied by a discontinuous decline in labor supply—across three data sets, which previous literature found difficult to explain. I propose and estimate a life-cycle labor supply model with habit persistence wherein sharp retirement can be explained by workers quitting “cold turkey.” In much the same way that one might quit smoking, workers with accumulated “working habit” exit the labor force with a pronounced, discontinuous decline in labor supply. The working habit model is consistent with the data, where workers reduce yearly labor supply by scaling back more in hours worked per week (over 50% reduction) than in weeks worked per year (20% reduction). The fixed costs approach, which has been the standard model used to understand sharp retirement, cannot explain these trends. After estimating the model, counterfactuals show that reducing Social Security benefits by 20% causes individuals work an additional 8.6 months. Individuals choosing sharp retirement respond mostly on the extensive margin by delaying retirement eight months, while individuals choosing smooth retirement respond mostly on the intensive margin by increasing yearly labor supply and delaying retirement only one month.

KEYWORDS: Retirement behavior, habit persistence, working habit, cold turkey

JEL Classification: J22, J26

1. Introduction

Across all three widely used data sets—the Current Population Survey (CPS), the Panel Survey of Income Dynamics (PSID) and the Health and Retirement Survey (HRS)—over 80% of individuals retire “sharply,” accompanied by a discontinuous decline in labor supply. In contrast to previous

*Office G22G, Ground Floor, East Lobby, ASB Building, UNSW, Sydney, NSW 2052, Australia. Email: x.fan@unsw.edu.au. I thank John Kennan for invaluable guidance and discussion. I also thank Ananth Seshadri and Christopher Taber for their support throughout the development of this paper. I am grateful to Steven Durlauf, Chao Fu, Jean-Francois Houde, Michael Keane, Sang Yoon Lee, Rasmus Lentz, Francesco Lippi, Rodolfo Manuelli, Salvador Navarro, Nicolas Roys, John Rust, Carl Sanders, James Walker, Ping Wang, and the summer labor group participants for helpful discussions and advice. I received help from Maria Casanova in understanding earnings test rules. All comments are welcome and all errors are solely mine. Please visit <https://sites.google.com/site/fanxiaodong/research> for latest version.

studies which use fixed costs to explain this phenomenon, I propose and estimate a working habit model and demonstrate that it delivers more empirically plausible results. Specifically, the working habit model predicts that workers would maintain a certain level of working habit by spreading time at work throughout an entire period, while the fixed costs model implies that workers would choose to concentrate their time at work in order to reduce the number of times they have to work. In the data I find when workers reduce yearly labor supply they scale back more in hours worked per week (over 50% reduction) than in weeks worked per year (20% reduction), which favors the working habit model.

I begin by documenting retirement behavior of white male workers in three data sets: CPS, PSID and HRS. Two different retirement paths emerge. Some workers first gradually reduce their labor supplies as they age and then retire, while other workers are never observed working part-time before retiring. In other words, this group of workers is observed retiring directly from their full-time jobs and quitting the labor market abruptly. I use the term “smooth retirement” to describe retirement behavior which smoothly reduces labor supply before retirement and the term “sharp retirement” for abrupt retirement behavior—completely retiring from full-time working. Over 80% of retirement is sharp in all three data sets.

In a neoclassical life-cycle model with intertemporally separable, concave utility in labor supply, individuals smooth labor supply (or leisure) across time.¹ This leads to a gradually declining labor supply profile as productivity decreases gradually with age, thus predicting smooth retirement. The abrupt drops in labor supply associated with sharp retirement are hard to explain with this model. To fill the gap, two not unrelated approaches are widely used in the literature: discrete labor choices or fixed costs of working. Discrete labor choice models assume that individuals can only choose to work full-time or not at all. Such labor market rigidity could be a result of fixed costs or team production spillover on the demand side, or fixed costs or firm-specific human capital on the supply side, or both. This approach could be seen as a reduced form application of the fixed costs approach, which assumes that individuals have to pay fixed costs, either money or time, to work.² This generates an incentive to cluster labor supply. There is empirical evidence of such fixed costs. In the PSID data a full-time worker pays 0.64 hours per day on average to commute between home and work. However, this is far from enough to explain sharp retirement. For instance, [French \(2005\)](#), [Rogerson and Wallenius \(2010\)](#), [French and Jones \(2011\)](#) estimate that fixed costs as four hours or higher per day are needed to generate sharp retirement.

I propose an alternative approach—habit persistence—to explain sharp retirement, which has a better fit to the data. In a life-cycle labor supply model with working habit, individuals accumulate working habit at work, which leads to sharp retirement. This is because working habit makes the worker more comfortable working by reducing the marginal disutility of working, due to complementarity of labor supply among adjacent periods. This adjacent complementarity generates an incentive to cluster labor supply. When such an incentive dominates the incentive to smooth

¹Or age. In this paper only one cohort is investigated so *time*, *age* or *period* are used interchangeably.

²This paper mainly focuses on the supply side. If one assumes that utility is transferable between the employer and the employee then the analysis on the supply side can be easily transferred to the demand side with similar results.

labor supply because of the convex disutility of working, it is optimal to stop working abruptly from a relatively high level of hours worked. In other words, while one individual is working he is in a “working mode.”³ If he plans to quit working and retire, he wants to get out of the working mode as quickly as possible. Gradually reducing labor supply is not optimal because of the nonconcavity of the value function that the adjacent complementarity of labor supply generates.

This kind of adjacent complementarity is initially studied in [Ryder and Heal \(1973\)](#) and is further developed in [Becker and Murphy \(1988\)](#) where they refer to it as rational addiction. In a quadratic utility example with two steady states [Becker and Murphy \(1988\)](#) show a bang-bang solution is optimal. Their paper did not focus on the dynamics which might generate cold turkey quitting, while these dynamics are the main ingredient of this current paper.

The idea of adjacent complementarity of labor supply is also closely related to the literature of intertemporally nonseparable preferences in the labor supply or leisure. Even though most empirical research uses intertemporally separable preferences for convenience or simplicity, intertemporal nonseparabilities in labor supply or leisure have repeatedly been shown to be significant.⁴ All the previous literature listed studies the intertemporal nonseparability on the intensive margin, while this paper applies it to the extensive margin to explain sharp retirement.

The comparison between the working habit model and the fixed costs model shows that the data appears more favorable to the working habit approach than the fixed costs approach. The fixed costs model predicts that workers would cluster labor supply within a period to minimize fixed costs. If one period is defined as one year and one assumes fixed costs are paid each week at working, then within one period an individual would try to cluster weeks worked per year but not hours worked per week.⁵ In contrast, the working habit model implies workers want to smooth labor supply within the period to maintain a certain level of working habit. In the PSID data and the HRS data, I find individuals who reduce yearly labor supply are more likely to reduce hours worked per week (from around 45 to below 20—a more than 50% deduction) rather than weeks worked per year (from 50 to 40—a 20% reduction), which fits better with the working habit model than the fixed costs model. This implication of the working habit model also helps justify the work-sharing program currently provided by the Employment Development Department (EDD) of California and implemented in Germany.

I use the method of simulated moments to match life cycle profiles estimated using the HRS data. I allow for heterogeneous preferences in the degree of adjacent complementarity to accommodate both sharp retirement and smooth retirement. I match life-cycle profiles of assets, labor supply, labor force participation rates, wages, and frequencies of sharp and smooth retirements on the population level, as well as conditional assets, labor supply and wage profiles for those

³This paper focuses on explaining sharp retirement of male workers so the pronoun “he” is used throughout. Female labor supply and retirement involve different complexities and are out of the scope of this paper. Interestingly the retirement pattern of female workers is very similar as discussed in the current paper.

⁴[Hotz et al. \(1988\)](#), [Bover \(1991\)](#), [Altug and Miller \(1998\)](#), [Woittiez and Kapteyn \(1998\)](#), [Kubin and Prinz \(2002\)](#).

⁵If fixed costs are paid each day, this is equivalent to that the yearly fixed costs are proportional to yearly labor supply. That is, in such a setup fixed costs are equivalent to a deduction in wage, and the model is similar to the standard neoclassical model which finds difficulty in justifying sharp retirement.

choosing sharp retirement. The wage profile parameters are estimated jointly with the structural model. The dynamic programming model generates reasonable parameter estimates with good model fits.

Understanding how individuals retire, or why individuals retire sharply, is important in two aspects. First, it helps us model the real world better. Second, it is critical in conducting counterfactual experiments. For example, when estimating the effects of changes in Social Security policy rules on lifetime labor supply and retirement behavior, it could make a big difference whether sharp retirement is an optimal outcome given unconstrained, continuous labor choices, or a result of labor market rigidity with constrained, indivisible labor choices. When the policy or environment changes, individuals behave differently in these two different models. Indeed, I find individuals choosing different types of retirement respond differently to changes of the Social Security benefits. Overall, reducing Social Security benefits by 20% makes an average individual work an additional 8.6 months, regardless of retirement type. However, individuals choosing sharp retirement respond mostly on the extensive margin by delaying retirement eight months, while individuals choosing smooth retirement respond mostly on the intensive margin by increasing yearly labor supply and delaying retirement only one month. In contrast, a model with labor market rigidity would predict that individuals have to respond on the extensive margin by delaying retirement in the absence of adjustment on the intensive margin.

In addition to changing the generosity of Social Security benefits, I use the model to conduct another two counterfactual experiments based on the estimated parameters. First, I shift the Early Retirement Age (ERA) from 62 to 64. Second I consider eliminating the Social Security earnings test. I find that increasing the ERA or eliminating Social Security earnings test has moderate or little effect on labor supply or retirement ages, while reducing the generosity of Social Security benefits universally has much larger effects on the labor supply and retirement ages due to the income effect.

The remainder of the paper is organized as follows. Section 2 documents retirement behavior in three data sets and reveals the dominance of sharp retirement. Section 3 develops the life-cycle labor supply model with working habit. Section 4 describes the data and estimation strategies. Section 5 presents parameter estimates. Section 6 reviews previous approaches explaining sharp retirement, followed by detailed comparisons in Section 7. Section 8 conducts three counterfactual experiments and Section 9 concludes.

2. Retirement in the Data

In this section, I document the retirement behavior of white male workers in three different data sets: the Current Population Survey (CPS), the Panel Survey of Income Dynamics (PSID), and the Health and Retirement Survey (HRS). I examine the dynamics of workers' labor supply in the transition to retirement, and find that some workers smooth their labor supply on the path to retirement while most do not.

I merge the CPS "Merged Outgoing Rotation Groups" (MORG) with the CPS "Annual De-

mographic File” (March) to get a short panel with four data points for each individual: Year Y (March), Year $Y + 1$ (March-June), Year $Y + 1$ (March) and Year $Y + 2$ (March-June). The PSID sample includes data between 1968 and 1997 when the survey is conducted yearly, which allows for better observation of employment-to-retirement transitions. The HRS sample includes only the initial HRS cohort (born 1931 to 1941, aged 51-61 as of the first wave in 1992, followed through nine waves until 2008 when respondents were aged 67 – 77). In all three data sets, only white male workers are included, so as to get a more homogeneous sample. For further details regarding data construction, please refer to Appendix A.

One of the main findings is that across all three data sets, most male workers are never observed working part-time before retiring. Over 80% of male workers are observed retiring directly from their full-time jobs, quitting the labor market abruptly.

From the merged CPS data, I choose a subset of white males who fully retired exactly in the last data point. Figure 1 plots the histogram of workers’ yearly and weekly labor supply in the year prior to their retirement, or the data point 3.⁶ The average yearly labor supply is 1,363 hours with a big spike at 2,100 hours and a smaller one around 1,100 hours. 52.5% of those individuals work for more than 1,260 hours, which is defined as full-time, in the year prior to their complete retirement.⁷ If the full-time is defined as $35 \times 45 = 1,575$ or more hours a year, then the ratio is slightly smaller, 42.5%. The ratios barely vary across marital status, work type, or health status (Table 2).

Yearly labor supply immediately before retirement could be misleading, since workers can retire at any time in a given year. The weekly labor supply is more concentrated around 40 hours. Between 0 and 40 hours, there is a small peak around 20 hours and the rest is almost uniformly distributed. Compared to the yearly level, overall a higher percentage of workers work full-time at the weekly level, 75.1% or 89.1% depending on whether full-time is defined as 35+ hours or 20+ hours.

For simplicity of description, I define “smooth retirement” as the retirement behavior whereby workers smoothly reduce their labor supply before retirement and “sharp retirement” as the behavior whereby workers retire abruptly from full-time jobs.

To check whether those who work part-time prior to retirement have indeed reduced their labor supply from previous full-time jobs, Figure 2 plots the labor supply profiles for individuals who choose sharp versus smooth retirement. The labor supply of workers choosing smooth retirement is gradually reduced over the two-year span at CPS. The average weekly labor supply for those workers is 28 hours two years prior to retirement and drops to only 18 hours a year later. This monotone reduction in labor supply over time differs from the profile of prime-age workers, where such a reduction of labor supply appears to be random. Figure 3 illustrates that more than half of workers working part-time at year $y + 1$ (March-June) are working full-time in

⁶The third time they are observed in the CPS data. Please refer to Appendix A.1 for detailed explanation of data points.

⁷Rand version HRS defines full time work as 35+ hours per week for 36+ weeks. I slightly relax it and define full time work as $35 \times 36 = 1,260$ hours per year. It didn’t make much difference in the data.

the observations immediately before or after. In the histogram for prime-age workers (Figure 4) labor supplies are almost all concentrated between 2,000 and 3,000 hours per year or 40+ hours per week. Part-time jobs are very rare for this group of workers, only around 3% in a given time.

A similar pattern is found in the PSID data. The yearly labor supply distribution indeed shifts toward part-time in the two years before retirement (Figure 5). At the yearly level, 35.8% of workers continue to work full-time the year before they completely retire from the labor market,⁸ which likely underestimates the prevalence of sharp retirement due to the mid-year retirement. Indeed at the week level, the ratio of full-time workers goes up to 82.1% (Table 2). The dynamics of labor supply reveal that individuals choosing smooth retirement gradually cut back their hours worked per year before full retirement (Figure 6).

Findings from HRS data again confirm that the majority of workers are never observed reducing labor supply before retirement. Two years prior to full retirement, over 80% of individuals are still working full time (Figure 7). The ratio is nearly the same at the weekly level (Table 2).⁹ Figure 8-9 reveal that, for individuals choosing smooth retirement, the gradual reduction of yearly labor supply comes mainly from the cutback in weekly labor supply. The hours worked per week decrease gradually from over 40 down to 20—more than 50% reduction—over ten years while the weeks worked per year decrease only from 50 down to 40—a 20% reduction—before retirement.

Table 2 summarizes the ratios of individuals working full-time at different interview points before retirement. Many individuals are never observed working part-time before retirement, regardless of marital status, work type, or health status; even among self-employed workers, 30%-75% of them choose sharp retirement.

The findings in this paper are consistent with those in the literature. Rogerson and Wallenius (2010) document the cross-sectional distribution of labor supply by age in pooled CPS data and PSID data. Blau and Shvydko (2011) report the wave-to-wave transitions of labor force status in HRS data. Both papers conclude that sharp retirement prevails.

Sharp retirement is difficult to explain in a neoclassical life-cycle labor supply model without assuming abrupt changes in preferences or constraints. In order to explain this puzzling retirement behavior, in addition to institutional restrictions such as Social Security or pension rules, two explanations are widely used in the literature: discrete labor choices or fixed costs of working. In this paper, I propose an alternative approach—a working habit model—to explain sharp retirement, presented and estimated in the next three sections.

3. Model

I propose and estimate a life-cycle labor supply model with working habit wherein sharp retirement can be explained by workers quitting “cold turkey.”

⁸Full-time is defined as 1,260 hours or more per year if not specified.

⁹Since it is conducted biennially, mid-year retirement is not a big issue in HRS.

3.1 The Set-up

This section describes a life-cycle working habit model where individuals make decisions regarding consumption, labor supply, and Social Security benefits application. The household head is considered to be the one who makes all relevant decisions. There are two sources of uncertainty each individual faces: survival probability and stochastic wages. Retirement is defined as a labor supply of zero.

Time is discrete. At time t , each living individual derives utility from consumption and labor supply,

$$U(c_t, h_t, B_t) = \frac{c_t^{1-\rho}}{1-\rho} - \alpha_h \frac{h_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} + \alpha_{hs} h_t S_t - \alpha_s S_t, \quad (1)$$

with $\alpha_h > 0$, $\alpha_{hs} > 0$, $\alpha_s > 0$.¹⁰ In addition to choosing consumption c_t and labor supply h_t , each individual, if eligible, decides whether to apply for Social Security benefits or not. B_t equals one if the individual decides to apply for the Social Security benefits and is zero otherwise.¹¹ While working, each individual accumulates a stock of working habit, measured by S_t , according to the following process,¹²

$$S_{t+1} = (1 - \delta)S_t + h_t. \quad (2)$$

The introduction of working habit in the preference is essential in this model, $\alpha_{hs} h_t S_t - \alpha_s S_t$, and will be discussed in greater detail in Section 3.3.

The individual faces a budget constraint

$$A_{t+1} = \begin{cases} (1+r)A_t + Y_t(w_t(h_t - \tau), y_t^f) + ss_t - c_t + tr_t - M, & \text{if } h_t > 0 \\ (1+r)A_t + Y_t(0, y_t^f) + ss_t - c_t + tr_t & \text{if } h_t = 0 \end{cases}, \quad (3)$$

where A_t is assets and r is the risk free interest rate. $Y_t(\cdot)$ is after-tax income, which is a function of wage income $w_t(h_t - \tau)$, other income y_t^f , and the tax code.¹³ ss_t is the Social Security benefits and tr_t is government transfer. If the individual works, the time fixed costs, τ , and the money fixed costs, M , have to be paid. These fixed costs are assumed to be zero in the baseline model estimated in Section 5. They are included here to demonstrate a general model, as well as serving a comparison between the working habit model and the fixed costs model in Section 6.3 and 7.1.

For a household with more than one member, the adult-equivalent consumption is calculated according to the formula

$$\frac{\hat{c}_t}{c_t} = (\#adults + a_1 \times \#kids)^{a_2} \quad (4)$$

¹⁰Subscript i is suppressed to save notation.

¹¹Assume a qualified application for Social Security benefits never gets denied.

¹²The way that an individual accumulates working habit is essentially similar as the learning-by-doing process estimated in Imai and Keane (2004). The difference is that working habit affects the utility directly while learning-by-doing affects the wage process directly. However, the learning-by-doing model predicts smooth retirement, not sharp retirement. Given the wage process and the retirement behavior, these two mechanisms could be separately identified.

¹³Please refer to Appendix B for the detailed tax code.

where $\#adults$ is the number of adults and $\#kids$ is the number of children in the household. \hat{c}_t is total consumption. Values of $a_1 = 0.70$ and $a_2 = 0.75$ are used in the paper following [Citro and Michael \(1995\)](#), page 178.

Individuals are not allowed to borrow against future income. This implies a non-negative borrowing constraint,

$$A_{t+1} \geq 0 \quad (5)$$

At time t , individuals face survival uncertainty. Conditional on being alive at time t , the probability of being alive at time $t + 1$ is $q_s(t + 1)$, which is a function of age and the Average Indexed Monthly Earnings (AIME).

Each individual also faces a second form of uncertainty, a stochastic wage process. The logarithm of wages at time t , $\ln w_t$, is a function of age plus an autoregressive component ζ_t ,

$$\ln w_t = x_t^w \pi_w + \zeta_t \quad (6)$$

The autoregressive component has a correlation coefficient ϕ_w and an independent, identically distributed innovation ϵ_t^w drawn from a normal distribution,

$$\zeta_t = \phi_w \zeta_{t-1} + \epsilon_t^w, \epsilon_t^w \sim N(0, \sigma_w^2) \quad (7)$$

In addition to the labor income and Social Security benefits, if applicable, each household also receives other income y_t^f , which includes all other sources of income and expenses. In particular, for this paper, other income could include pension benefits, income from other household members, and medical expenses (so it could be negative). For the sake of computational simplicity, other income is assumed to be a deterministic function of age, AIME, and their interaction,

$$y_t^f = x_t^f \pi_f \quad (8)$$

where x_t^f includes age and its square, AIME and its square, and the interaction of age and AIME. Ignoring the complication of private pension rules reduces the computational burden at the cost of possible overestimation of the degree of adjacent complementarity. Similarly, ignoring stochastic medical expenses comes at a cost of overestimating the bequest motivation.

An individual may suffer large, negative other income from medical expenses, in which case government transfers provide a consumption floor ([Hubbard et al. \(1995\)](#))

$$tr_t = \max \{0, c_{min} - ((1 + r)A_t + Y_t + ss_t)\}. \quad (9)$$

When $tr_t > 0$, $c_t = c_{min}$ and $A_{t+1} = 0$.

There is a bequest motivation in the form of

$$b(A_t) = b_1 \frac{(b_2 + A_t)^{1-\rho}}{1-\rho} \quad (10)$$

where b_1 captures the relative weight of the bequest motivation and b_2 determines its curvature ([DeNardi \(2004\)](#)).

At each time t , each individual seeks (B_t, h_t, c_t) to maximize the present value of lifetime utilities, or the value function in a recursive manner,

$$V(X_t) = \max_{B_t, h_t, c_t} \frac{c_t^{1-\rho}}{1-\rho} - \alpha_h \frac{h_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} + \alpha_{hs} h_t S_t - \alpha_s S_t + \beta (1 - q_s(t)) b(A_{t+1}) + \beta q_s(t) \int V(X_{t+1}) dF(X_{t+1}|X_t, c_t, h_t, B_t). \quad (11)$$

subject to the law of motion for working habit (2), the budget constraint (3), the borrowing constraint (5), the wage process (6)-(7), the other income process (8), the government transfers 9, as well as the bequest motivation (10). X_t is the vector of state variables, $X_t = (A_t, S_t, AIME_t, \zeta_t, B_{t-1})$.

3.2 Model Solution

The model is solved numerically. At time t , the state space is discretized into a finite grid space: log-linearly into 10 points, respectively, in asset A_t and $AIME_t$, and linearly into 10 points in working habit S_t . The autoregressive component in the wage process is discretized into a discrete Markov process with 7 grids using the Rouwenhorst method (Kopecky and Suen (2010)). The Social Security benefits application B_t is already discrete, with two points. This results in a grid of 14,000 points in each period.

The value function and decision rules are solved backwards. Linear interpolation is used in calculating the next-period value function and policy functions at any state value. The state space is carefully set such that it does not involve extrapolation which is usually unstable.

At each period, the value function is solved as shown in (11). At any given grid point of state variables X_t , the value of taking Social Security benefits ($B_t = 1$) or not ($B_t = 0$) is calculated and the higher value is picked as optimal. In either case, the value is maximized over the labor supply h_t first where, for any h_t , the Euler equation of consumption (12) is utilized to solve the optimal consumption $c_t(X_t, h_t, B_t)$.

$$c_t^{-\rho} - \beta \left\{ (1 - q_s(t)) \frac{\partial b(A_{t+1})}{\partial A_{t+1}} + q_s(t) (1 + r) \int c_{t+1}^{-\rho} dF(X_{t+1}|X_t, c_t, h_t, B_t) \right\} \geq 0, \quad (12)$$

$$= \text{if } c_t < (1 + r)A_t + Y_t + sS_t$$

$$\geq \text{if } c_t = (1 + r)A_t + Y_t + sS_t.$$

3.3 Working Habit

Working habit enters the utility function (1) in two forms: reinforcement and tolerance. Reinforcement, reflected in the term $\alpha_{hs} h S$, is closely related to the idea of adjacent complementarity (Ryder and Heal (1973)): current labor supply reduces the marginal disutility of future labor supply. That is, the more one works today, the more one wants to work tomorrow. Another term in (1), $-\alpha_s S_t$, reflects tolerance; current labor supply lowers future utility for given future labor supply.

This reinforcement is just one form of intertemporal nonseparability in labor supply. Even though most research uses intertemporally separable preferences for convenience or simplicity, intertemporal nonseparability in labor supply has been tested in different data sets and almost all relevant research rejects the separability hypothesis. They show that past labor supply does affect current labor supply decisions, although they disagree on whether the past and current labor supplies are substitutes or complements. In a translog utility specification, [Hotz et al. \(1988\)](#) estimate past leisure as a substitute for current leisure in a linear form. Using a similar setup and data, [Bover \(1991\)](#) finds that past labor supply is a complement for current labor supply. These two papers use a specification where past labor supply (or leisure) interact in a linear form. A fair amount of other literature directly investigates a more general interaction of labor supply among adjacent periods. Results are mixed. Both intertemporal substitution ([Kennan \(1988\)](#), [Altug and Miller \(1998\)](#)) and complementarity ([Miller and Sanders \(1997\)](#), [Woittiez and Kapteyn \(1998\)](#)) are supported in data.

Note that while the research above estimates the intertemporal nonseparability in labor supply on the intensive margin, this paper applies and estimates intertemporal nonseparability on the extensive margin.

The tolerance term, $-\alpha_s S_t$, is interpreted as disutility of labor supply paid over time. It could be physical or mental disutility. The tolerance term is different from the fixed costs used in [Cogan \(1981\)](#), [French \(2005\)](#), [French and Jones \(2011\)](#), or [Rogerson and Wallenius \(2010\)](#). In this model the tolerance term is proportional to labor supply, while in a model with fixed costs, a fixed amount of costs has to be paid for any positive amount of labor supply.¹⁴ Cold turkey quitting is not generated by this tolerance term. Technically it just generates a positive marginal disutility of working at zero labor supply so that it implies a positive reservation wage.

These two properties (reinforcement and tolerance) are very similar to those of a harmful addiction, a concept fully developed in the seminal paper, [Becker and Murphy \(1988\)](#), although I am reluctant to compare working habit with harmful addictions. Many addictions, such as heavy smoking, gambling, heavy drinking, or over eating, are considered to be inherently bad and should be terminated whenever possible. Working habit (not workaholic), however, is typically considered a virtue.

Notice that the utility function (1) is concave in h or S individually, but not jointly in h and S . [Becker and Murphy \(1988\)](#) show that “cold turkey” is the only way to end a strong addiction. This feature makes this model different from canonical utilities and makes sharp retirement optimal. The idea is that the worker accumulates working habit at work. Such working habit makes the worker more and more comfortable (or less and less uncomfortable) working by reducing the marginal disutility of working. That is, labor supply is complementary among adjacent periods. Such adjacent complementarity generates an incentive to cluster labor supply. When this incentive dominates the incentive of smoothing labor supply, which comes from the convex disutility of

¹⁴It is also different from [Donald and Hamermesh \(2009\)](#). In their model participating the labor market reduces efficiency at home production or leisure but the degree of efficiency lost is independent of the amount of labor supply as long as it is positive.

working, it is optimal to stop working abruptly from a relatively high-hour level. In other words, while the worker is working he is in the working mode. If he plans to quit working and retire, he wants to get out of the working mode as quickly as possible. The nonconcavity of the value function generated from the adjacent complementary results in a discontinuous labor supply policy function.

An example illustrates the life-cycle labor supply profile in a working habit model. Figure 10 shows a simplified model with a linear wage profile from 5.0 in periods 30-34 to 0.0 at the last five periods. Assume no Social Security benefits. The model is solved for its initial values of $A_{30} = 0.0$ and $S_{30} = 0.4$. The life cycle labor supply profile is simulated. It is shown in the figure that the labor supply is highly inelastic on the intensive margin but extremely elastic on the extensive margin. The simulated individual tries to reduce labor supply as quickly as possible and retires sharply.

This utility function is consistent with another form of habit persistence—adjustment costs. A model with adjustment costs assumes that workers have to pay some adjustment cost if their current labor supply differs from previous levels,

$$U(c_t, h_t, S_t) = U(c_t, h_t) - (h_t - S_t)^2 = U(c_t, h_t) - h_t^2 + 2h_t S_t - S_t^2.$$

This also implies an adjacent complementarity on labor supply. The difference is that here the utility function is always jointly concave in h_t and S_t , which implies workers will smooth both.

Working habit can be interpreted as working routines. After spending many years working, many workers build up working routines which make working more pleasant than the case without such routines. A relevant example is that getting up early for work is usually a much harder job for younger workers than for older workers. Another example at the higher frequency is the weekly routine. Some people have “Monday blues.” This could be explained by that on Mondays their weekend routine breaks but the weekday working routine, which is broken by the weekend, has not been re-established yet. Through Mondays they again build up their weekday working routine which makes working on other weekdays more pleasant than on Mondays, therefore Mondays become the worst day of a week.¹⁵

4. Data, Calibration and Estimation

I use Rand version HRS data to estimate the parameters of interest. Please refer to Appendix A for sample selection criteria. I use the original HRS cohort where individuals were aged 51-61 when initially interviewed in 1992, after which data was collected biennially. This paper uses 9 waves of data, through 2008, when the cohort was aged 67-77.

I assume that each individual can live as long as 85 years. This seems arbitrary but reasons are twofold. First, this paper concerns retirement behavior, not retiree behavior. Of course the

¹⁵However, the working habit model presented in this paper is not able to explain why workers take weekends off. The allocation of working days (or hours) during the week (or the day) is a very interesting topic itself but is beyond the scope of this paper. Please refer to Footnote 36.

uncertainty and risk (health, medical expenses, et al) that each individual faces after retirement are important in shaping his retirement timing as well as his consumption and saving decisions through the precautionary saving motivation (for example, [DeNardi et al. \(2010\)](#)). For this reason, a more altruistic bequest motivation is expected in this paper than occurs in most literature. After controlling for such a precautionary saving effect, life expectancy should not affect retirement timing or behavior. Second, it is computationally advantageous to have fewer periods.

One period is defined as two years because hours worked are only observed every two years in the HRS data. As discussed previously, the two-year gap between adjacent interviews will almost certainly exaggerate sharp retirement behavior. For example if a smooth transition from full-time working to full retirement takes place within a two-year horizon, what appears to be a sharp retirement in the data may be indeed a smooth one. However, evidence from the CPS data and the PSID data, where data are collected each year, shows the prevalence of sharp retirement at the weekly or yearly level. Furthermore, for nearly a quarter of workers who choose smooth retirement, on average the smoothing process spans ten years ([Figure 8](#)). Compared to them, the retirement which reduces labor supply from full-time to zero within a two-year span is still relatively sharp. Thus defining two years as one period does generate some exaggeration, but is unlikely to be severe.

The model starts at period 50 and ends at period 67, which is equivalent to age 84-85 in the data. The early retirement age is at period 56 and the normal retirement age is at period 58.

The risk free interest rate r is set at 0.06. The consumption floor is set at 0.78.¹⁶

The vectors of parameters are listed below,

$$\begin{aligned} \Theta &= (\Theta_1, \Theta_2) \\ \Theta_1 &= \left\{ \underbrace{q_s(\cdot)}_{\text{survival}}, \underbrace{\pi_f}_{\text{other income}} \right\} \\ \Theta_2 &= \left\{ \beta, \underbrace{\rho}_c, \underbrace{\gamma, \alpha_h, \{\alpha_{hs}^j, \alpha_s^j\}_{j=1,2}}_{\text{labor supply}}, \underbrace{\kappa_2, \kappa_3, \kappa_4}_{\text{Prob}(j=1)}, \underbrace{\delta, \kappa_0, \kappa_1}_S, \underbrace{b_1, b_2}_{\text{bequest}}, \underbrace{\pi_w, \phi_w, \sigma_w^2}_{\text{wage}} \right\} \end{aligned}$$

The model is estimated in two steps. The first step is to estimate parameter set Θ_1 on the survival probabilities and the other income process from the data directly. The second step is to estimate all remaining parameters Θ_2 , including the wage process using the method of simulated moments (MSM).

4.1 Moment Conditions

The method of simulated moments is to find a set of parameters which is able to generate simulated life-cycle profiles “closest” to the data. Closeness is usually measured by the distance be-

¹⁶0.78 = 4,380/5,600 ([French and Jones \(2011\)](#)).

tween the moments from the simulated profiles and the same moments from data. In this paper the distance is measured by a weighted squared error.

Two sets of moment conditions at each age are chosen to represent the life-cycle profiles.

The first set of moment conditions includes:

- (i) Assets: the first and the second moments (mean and variance).
- (ii) Logarithm of wages: the first and the second moments.
- (iii) The labor force participation rate.
- (iv) Labor supply: the first and the second moments.
- (v) The ratio of individuals who choose sharp retirement versus smooth retirement to the whole population.
- (vi) The job re-entering rate.

The moment conditions of log wages and labor supply are calculated as follows

$$E((Z_{it} - \bar{Z}_t) \cdot D_{it}) = 0 \quad (13)$$

where $D_{it} = 1$ if an individual is working and $D_{it} = 0$ otherwise, $Z_{it} \in \{\log \text{ wage, labor supply}\}$. \bar{Z}_t is the model's prediction of Z_{it} conditional on $D_{it} = 1$

$$\bar{Z}_t = \frac{\sum \hat{Z}_{it}}{n_t} | (D_{it} = 1) \quad (14)$$

where \hat{Z}_{it} is the log wage, or labor supply in the simulated data, and n_t is the number of cases of $D_{it} = 1$.

The second set of moment conditions is for individuals choosing sharp retirement ignoring re-entry from retirement:

- (vii) Assets: the first and the second moments.
- (viii) Logarithm of wages: the first and the second moments.
- (IX) Labor supply: the first and the second moments.

The second set of moment conditions is calculated as follows

$$E((Z_{it} - \bar{Z}_t) \cdot D_{it}) = 0 \quad (15)$$

where $D_{it} = 1$ if an individual is working and will choose sharp retirement later during the sample period and $D_{it} = 0$ otherwise, $Z_{it} \in \{\text{assets, log wage, labor supply}\}$. \bar{Z}_t is the model's prediction of Z_{it} conditional on $D_{it} = 1$

$$\bar{Z}_t = \frac{\sum \hat{Z}_{it}}{n_t} | (D_{it} = 1) \quad (16)$$

where \hat{Z}_{it} is the assets, log wage, or labor supply in the simulated data, and n_t is the number of cases of $D_{it} = 1$.

From the data, between individuals who choose sharp retirement and those who choose smooth retirement, when they re-enter the labor force after retirement their labor supply and wages are similar. For this reason, I tailor the labor supply and wages at retirement in the second

set of moment conditions to emphasize the difference of labor supply and wages before retirement between these two groups of individuals.

The first period in the moment conditions is set to be period 52 since all period 51 profiles are initial conditions. The final period in the moment conditions is period 61. The job re-entering rates at period 52-53 are zero so these two moments are dropped as well. The number of individuals who choose sharp retirement but still work after period 60 (equivalent to age 70) is very small so they are also excluded. This leaves 146 moment conditions in total.

4.2 Initial Conditions and Preference Heterogeneity

2,000 individuals are simulated. For each simulated individual, the initial conditions are drawn jointly from the data, including age, asset, AIME and wage.

The unobserved initial value of working habit S is assumed to be a linear function of the labor supply at the first period, $S_0 = \kappa_0 + \kappa_1 \cdot h_0$. This is to assume that the initially observed hours worked is a good proxy of the accumulated level of working habit. I also tried assuming S_0 is drawn from a normal distribution, with very similar results.

The initial conditions are to capture some of the heterogeneity in the data which are not captured by the model. For instance, individuals with high initial wage will more likely have high wages in following periods since the unobservable component in the wage process follows an AR(1) process with some persistence (parameter to be estimated). This will hopefully capture the effect of education which is otherwise missing in this model.

I also control for unobservable differences across retirement types by introducing permanent preference heterogeneity, following [Heckman and Singer \(1984\)](#), [Keane and Wolpin \(1997\)](#), [French and Jones \(2011\)](#). There are assumed to be two types of permanent preference heterogeneity in terms of having different $\{\alpha_{hs}^j, \alpha_s^j\}$, $j = 1, 2$. For each individual the probability of having $\{\alpha_{hs}^1, \alpha_s^1\}$ is modeled as a logistic function of a subset of initial state variables: log wage ($\ln w_0$) and age (t),

$$\Pr(j = 1) = \frac{1}{1 + \exp(-(\kappa_2 + \kappa_3 \ln w_0 + \kappa_4 (t_0 - 50)))} \quad (17)$$

I also tried including different variables, for example assets, but they were not significant and results were similar.

4.3 Estimation Procedure

The parameters are estimated according to the following standard procedure of the method of simulated moments.

- (1) Calculate the moments from the data.
- (2) Estimate the data generating processes from the first step: the survival probability and the other income process.
- (3) Simulate individuals by drawing their initial conditions from the joint distribution of the data.

- (4) Generate stochastic shocks (the surviving probability and the AR(1) transition probability) for each period for each simulated individual.
- (5) Pick a set of parameters and solve the value function (11) as well as the decision rules.
- (6) After solving the decision rules, generate the life-cycle profile for each simulated individual given the initial condition.
- (7) After getting the life-cycle profiles for all simulated individuals, calculate the simulated moments.
- (8) Compute the distance between the simulated moments and the data moments.
- (9) Pick another set of parameters, repeat steps (5)-(8) until a minimum distance is found.

5. Baseline Results

The baseline model assumes no fixed costs, $\tau = M = 0$. The estimated parameters are listed in Table 1. The model fits well in most moments.

The bequest weight—164.134—is much larger than the 0.0223 estimated in French and Jones (2011). As discussed in Section 4, this is expected because the maximum age is only 85 versus 94 in that paper and medical expenses are not included in this model, either.

The model predicts high saving rates with a high bequest weight (Figure 15). This is consistent with the findings in Scholz et al. (2006). The lack of uncertainty in the current model is the main reason for not being able to match the second moment of assets well. In the model the only two sources of uncertainty are from the wage process and the survival probability. After retirement, the survival probability is the main source of uncertainty, which is not able to generate enough variation. Adding variation in medical expenses for retirees will help improve the fit (DeNardi et al. (2010)).

To further check the fit of the model, I compare different simulated profiles which are not directly compared with the data. Figure 16 plots the retirement age distributions of individuals who choose sharp and smooth retirement. The model fits the data fairly well.

Figure 17 plots the labor supply profiles relative to the first retirement wave. The model fits well in hours for individuals choosing smooth retirement. For individuals choosing sharp retirement, the model predicts a slightly decreasing trend before retirement. This is due to the fact that one period is defined as two years in current specification. The simulation shows one individual does slightly decrease his labor supply in the period before full retirement. When the model is forced to fit the sharp retirement ratios, it requires higher labor supply before retirement. Instead, when one period is defined as a single year, and the model tries to fit moments for every two years, the labor supply fitting is much improved.

6. Previous Approaches

This section reviews how the previous literature models retirement and reconciles retirement behavior documented above, as well as their limitations.

Table 1: Parameters estimated in the baseline model.

Parameters		Estimates	Parameters		Estimates	
Habit	$\alpha_{hs,1}$	1.979 (0.0009)	Labor supply	γ	0.253 (0.0007)	
	$\alpha_{s,1}$	1.283 (0.0015)			α_h	13.927 (0.0021)
	$\alpha_{hs,2}$	0.013 (0.0617)	Time discount	β	0.985 (0.0004)	
	$\alpha_{s,2}$	0.517 (0.0029)			CRRA	ρ
	δ	0.907 (0.0010)	Wage: AR(1)	ϕ_w	0.890 (0.0003)	
Prob($j = 1$)		Wage: error	σ_w	0.307 (0.0004)		
cons	κ_2	0.821 (0.0057)	Wage: <i>age</i>	π_w	$4.597e - 3$ (0.0059)	
($\ln w_0$)	κ_3	0.010 (0.032)	Wage: <i>age</i> ²		$-4.561e - 4$ (0.0035)	
($t_0 - 50$)	κ_4	-0.010 (0.0031)	Wage: <i>cons</i>		3.185 (0.0006)	
Initial S	κ_0	0.277 (0.0011)	Bequest weight	b_1	164.134 (0.0021)	
	κ_1	0.231 (0.0013)	Bequest shifter	b_2	7.675 (0.0046)	

The standard neoclassical life-cycle labor supply model assumes $\alpha_{hs} = \alpha_s = 0$ in the utility (1), and $\tau = M = 0$ in the budget constraint (3). For simplicity assume no tax, no social security, no uncertainty in the wage process, and each individual lives to the last period.

The interior solution derived from the first order condition is

$$-\frac{U'_h(c_t, h_t)}{U'_c(c_t, h_t)} \geq w_t, \text{ if } h_t > 0$$

The worker stops working when the marginal cost of labor supply $-U'_h(c_t, h_t)$ exceeds the marginal benefit $w_t U'_c(c_t, h_t)$, at the corner $h_t = 0$. That is, $h_t = 0$ as long as $w_t \leq -U'_h(c_t, 0) / U'_c(c_t, 0) = w_{R,t}$.¹⁷ Curve A in Figure 20 illustrates the labor supply as a function of wage at time t with \hat{h} being the total time endowment. Workers retire whenever the reservation wage $w_{R,t}$ exceeds the after-tax wage w_t , which can occur due to an increase in the reservation wage, a decrease in the wage, or both.

The reservation wage could increase if the marginal utility of leisure or the marginal disutility of labor increases with age, $-\partial^2 U(c_t, h_t) / \partial h \partial t > 0$. Gustman and Steinmeier (1986) present a structural retirement model where workers are heterogeneous in terms of preference, face no uncertainty, and the utility of leisure is increasing in age. Gustman and Steinmeier (2006, 2008) apply similar models to estimate how Social Security policies affect retirement. Blau (2008) also assumes that the disutility of labor market participation increases with age. In all cases changes in the preference are assumed to be gradual.

Many factors can cause a decrease in after-tax wages. For example, in an optimal wage contract designed to solve the agency problem, Lazear (1979) shows the wage profile is backloaded; wages are lower than productivity when young but higher than productivity when old. After the

¹⁷In this specific example $w_{R,t} = 0$. Adding additional assumption can easily generate $w_{R,t} > 0$, for example $-\alpha_h \frac{h_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - \alpha_{h0} h_t$, or $-\alpha_h \frac{(h_t + \hat{h})^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$.

termination of such a backloaded contract, the wage falls back to the productivity level. If the decrease is large and sharp enough, retirement might occur. It is true that workers who did not switch jobs within the sample period are more likely to retire sharply (89.1% at the yearly level or 86.3% at the weekly level in HRS data), but sharp retirement also dominates for those who switched jobs within the sample period (58.1% at the yearly level or 52.8% at the weekly level in HRS data). Lazear (1979) assumes labor supply is a binary choice. It is not clear what an optimal contract would be if labor supply is allowed to be continuous, and it is not clear either how large the gap between wages and productivity for elderly workers would be. For this reason, this paper abstracts from the wage contract argument and assumes that one worker's pre-tax wage is equal to his productivity at every point in time.

The after-tax wages would decrease if productivity declines with age. In most cases, changes in productivity are also assumed to be gradual. Even though an abrupt decline in productivity can be caused by a bad health shock, Blau and Shvydko (2011) use HRS data and find that most retirement is not associated with such deterioration in health. Table 2 also shows sharp retirement dominates regardless of health status.

The decrease in after-tax wages can also be caused by Social Security tax policy. For example, the Social Security earnings test can reduce the after-tax wage dramatically. The earnings test, however, is just a delayed payment of Social Security benefits. It is actuarially fair for individuals aged 62 to 65 and it is close to being actuarially fair for most individuals at age 65 in the HRS cohort.¹⁸ Since 2000, the earnings test is eliminated after reaching the Normal Retirement Age (NRA). The only group of individuals which might retire sharply due to the earnings test is comprised of those who retire at age 65 or older before 2000. In the HRS, this involves about 6.5% of retirements and 64.7% of them are sharp, which is below the sample average.

Institutional rules, such as Social Security rules and private pensions, on the other hand, are likely able to cause discontinuity on Social Security or pension income at certain ages, thus they might be able to explain some of sharp retirement at those ages. However the effect is not large and there is still much sharp retirement behavior which remains unexplained, as discussed below.

6.1 Social Security or Pension Rules

The Social Security policy rules require that individuals are at least 62 years old to receive any benefits, and at least 65 years old to receive full benefit, as well as Medicare.¹⁹ In a model without saving, Rust and Phelan (1997) argue that individuals with low incomes would have to work

¹⁸In 2000, the life expectancy of those at age 65 was 17.9 years for the white population and 16.3 years for white males (National Vital Statistics Reports, Volume 51, Number 3). If a whole year's worth of benefits is withheld between ages 62 to 65, benefits in the future will be raised by 6.7% each year. If the benefit is withheld between age 66 to 70, future benefits will be raised by the amount shown in Table 13. These are close to being actuarially fair. French (2005) shows that removing earnings test workers would delay retirement almost one year, but he uses the PSID data for the years 1968–1997. This is a much older cohort than the data used in this paper. Thus the life expectancy is shorter and the future benefits are raised much less (Table 13), so the earnings test discourages working for individuals aged 65 or older in his data sample.

¹⁹Please refer to Section C for details on Social Security rules.

hard to finance consumption up until age 62 (or 65) to be qualified for Social Security benefits (or Medicare). However in a richer model with savings, French (2005) runs a simulation of shifting the early retirement age from 62 to 63 and finds that this has almost no effect on labor supply since most workers have savings at the retirement age. This implies that it is unlikely that Social Security rules keep individuals working full-time in the labor market until retirement age. Furthermore, sharp retirement could happen at any age (Figure 18).²⁰ Similar retirement age distributions are found in PSID and HRS data. It seems clear that the Social Security policy is not enough to explain sharp retirement.

Further, private pensions are only able to explain a small portion of sharp retirement. There are generally two different types of private pension plans: defined benefit (DB) pensions, which are closely attached to earning histories, tenures, and ages, and defined contribution (DC) pensions which depend mostly on the accumulated contributions of employers and employees and gains or losses from account investments. There are two possible channels through which these private pensions could affect an individual's retirement behavior. The first is the age restriction. Similar to Social Security, pension benefits are illiquid before a certain age (usually 55, 60 or 62), otherwise a strong benefit-deduction penalty applies. As discussed above in the Social Security case, for individuals who are not financially constrained such an age restriction does not have a significant effect on labor supply decisions. The second channel is through pension construction. DB pension benefits depend on one's tenure and highest annual earnings (usually the average of the five highest-earnings years) with a specific employer. DC pension benefits mainly depend on the accumulated account and is not directly associated with any specific employer. As French (2005) discusses, the construction of pension benefits generates incentives which govern an individual's retirement-timing decision. However, in either case, a pension does not typically restrict how much one has to work, thus does not necessarily imply sharp retirement.²¹ In HRS data, around 70% of individuals receive pension benefits. Among them, around 40% of retirements are associated with receiving private pension benefits, and over 95% of those retirements are sharp. This percentage is higher than the average and is consistent with the argument that the construction of private pension benefits might induce sharp retirement. On the other hand, among the other 60% of retirements, about 65% of them are also sharp, which cannot be explained by the pension system.

In sum, Social Security and private pension rules might be able to explain some of the sharp retirements in the data, but clearly unexplained factors remain. Outside the scope of such institutional rules, the model above implies that one individual would gradually reduce labor supply, implying smooth retirement.²² To explain sharp retirement, two closely related approaches are

²⁰Figure 18 has two spikes in retirement age before 2000 and only one after 2000. The disappearing spike at age 65 after 2000 could be explained by the elimination of the earnings test for workers aged 65 or older in that year. It is also true that at age 62, the ratio of sharp retirement is disproportionately high where the Social Security policy is likely to be the driver.

²¹Some employers might require a minimum of 1,000 hours per year to be covered under pension plans.

²²Mandatory retirement is prohibited in most jobs since 1994. The original Age Discrimination in Employment Act of 1967 (ADEA) "prohibits employment discrimination against persons 40 years of age or older", except "where age

used in the literature, namely, discrete labor choices and fixed costs.

6.2 Indivisible Labor

Some research assumes that labor choices are discrete (Rust and Phelan (1997); Casanova (2010)). Workers are only offered three choices: working full-time (40 hours per week), working part-time (20 hours per week), or not at all (0 hours per week).

Discrete labor choices could be caused by labor market rigidity, either from the demand side or from the supply side. Hurd (1996) reviews a rather complete list of possible sources of labor market rigidity. On the demand side, employment costs and team production spillovers might prevent employers from offering part-time jobs. On the supply side, if workers want to work part-time they might have to switch to low-skilled jobs where part-time workers are more desired. By doing so, workers might lose most or all job-specific skills. The competition for these low-skilled jobs is also more intense. In addition, workers with pre-existing health conditions might find it difficult to get similar health insurance coverage in a new job. Blau and Shvydko (2011) also list many factors which could make the labor market rigid.

Evidence from data suggests that the labor market does have some rigidity. One example listed in Hurd (1996) is that the ratio of self-employed workers observed in the data increases with age, given that self-employed workers have more flexibility in supplying labor. This is also confirmed in the CPS, PSID and HRS data where government and private workers are more likely to choose sharp retirement than are self-employed workers (Table 2). Workers switching occupations are indeed less likely to retire sharply than those who do not (Table 4).²³

The rigidity in the labor market could explain sharp retirement behavior, but it is unlikely to be the only determinant. There are some sharp retirements for which labor market rigidity is of limited explanatory power. Looking at most recent jobs, the ratios of sharp retirement are quite high in all occupations and industries (“All” in Figure 19). For self-employed workers, who are likely more flexible in their labor supply choices, it is surprising that nearly half of them choose sharp retirement. It is not clear either whether workers switch jobs voluntarily or due to rigidity. For those workers who do switch jobs, sharp retirement at their new jobs still dominates in many occupations and industries (“Switched jobs” in Figure 19).

6.3 Fixed Costs

Another strand of literature on retirement assumes that the individual has to pay some fixed costs in order to work—money costs or time costs. The effect of fixed costs on labor supply in a static setup is clearly demonstrated in Cogan (1981). Fixed costs of work induce workers to cluster their

is a *bona fide* occupational qualification reasonably necessary to the normal operation of the particular business” such as pilots and bus drivers. Mandatory retirement was eliminated in 1967 for workers aged 40-70 and the upper bound was removed in the amendment of 1986, with a few exemptions. Colleges and universities were temporarily allowed to enforce the mandatory retirement of age 70 until the beginning of 1994.

²³Percentages in the diagonal (skip self employment) are higher than those off the diagonal.

labor supply. Rogerson and Wallenius (2010) also demonstrate that fixed costs can generate sharp retirement even if productivity or preference does not change.

In a model with fixed costs, $\alpha_{hs} = \alpha_s = 0$, but $\tau > 0$ or $M > 0$ or both. The existence of fixed costs shifts the budget constraint of workers downward as shown in Figure 20.²⁴ Curve U_F denotes an individual's indifference curve in a model with fixed costs. Figure 20 shows the case with reservation wage $w_{F,t}$ and reservation hours $h_{F,t}$. Clearly, given wage $w_{F,t}$ any number of hours worked between 0 and $h_{F,t}$ generates less utility than the non-working case. Curve F designates the labor supply policy function. For high enough fixed costs,²⁵ the reservation labor supply will approach full-time work. In this case, an individual either works full-time for a wage above $w_{F,t}$ or does not work at all for wages below $w_{F,t}$ but he will never work part-time. This implies retirement must be sharp.

The model with fixed costs is one special case of a more general approach which assumes a (partially) convex mapping between labor supply and productivity (Prescott et al. (2009)). Such convexity may induce the worker to cluster labor supply instead of smoothing it.

It is clear that individuals need to pay certain costs to work. Time and money (gas, toll) spent to commute between home and work is probably the clearest fixed costs. Information collected in the PSID shows that the average commute time represents around 8% of labor supply—0.64 hours per day for a full-time worker.²⁶

However, such fixed costs are far from enough to generate sharp retirement. French (2005) estimates that in models without a convex wage-hours mapping²⁷ fixed time costs are 26%-29% of total leisure endowment, or 53%-59% of a full-time job,²⁸ which is 4.23-4.7 hours on top of any hours worked in each working day. Even including the convex mapping,²⁹ which is effectively equivalent to fixed costs, fixed time costs are 7%-9% of total leisure endowment, or 14%-17% of a full-time job, or 1.13-1.37 hours on top of any amount of hours worked in each working day. French (2005) calculates that a 62-year-old worker would never work less than 885 hours per year in the baseline model. This matches findings in Cogan (1981) where the estimated average annual fixed monetary cost of work is 28.3% of annual earnings for working, married women, which imparts average annual reservation hours of 1,288 hours, or 25.76 hours per week for 50 weeks. In French and Jones (2011), the estimated daily fixed time costs are 3.26 hours at age 60 and increase by 0.22 hours per year. By age 70, the fixed time costs are 5.46 hours per working day. In a demonstration model with constant wage, Rogerson and Wallenius (2010) calculate that

²⁴Taken from Figure 1 on Cogan (1981).

²⁵Strictly speaking, for high enough money costs. The effect of an increase in time costs on reservation hours is ambiguous. On one hand, the increase in time costs pushes up the reservation hours, given the total available time fixed. However, the increase in time costs also decreases the total available time allocated between labor supply and leisure.

²⁶This is the author's calculation, which matches the calculation in Gonzalez (2008) where the average commuting time is about 3 hours per week based on American Time Use Survey 2003.

²⁷Table 2, specification (1) and (2), without accounting for tied wage-hours offers.

²⁸Assume a full-time job is $(8 \times 5) / (16 \times 5) = 0.5$ of total leisure endowment. If one assumes the ratio to be $(8 \times 5) / (16 \times 7) = 0.36$, then the fixed time costs will be 74%-82% of a full-time job.

²⁹Table 2, specification (3) and (4), accounting for tied wage-hours offers.

fixed time costs need to be higher than a part-time job in order to create sharp retirement when the intertemporal elasticity of substitution (IES) of labor supply is 0.5 or lower. Thus in order to explain sharp retirement, fixed costs need to be implausibly high. This implies there could be other factors driving sharp retirement.

Models with fixed costs also fail to explain the fact in the data when workers scale back labor supply they scale back more in the hours worked per week than in the weeks worked per year. This will be discussed in greater detail in the following comparison section.

Another type of fixed costs is the fixed adjustment costs proposed in [Chetty \(2012\)](#). A fixed amount of costs has to be paid whenever the labor supply choice deviates from previous level. [Chetty \(2012\)](#) shows that a small amount of such friction is sufficient to reconcile the small labor supply elasticity on the intensive margin. Quantitatively it needs further work to estimate how much fixed adjustment costs are required to induce an individual to adjust labor supply only once (from full time to retirement) when wages decline. Intuitively such costs need to be sufficiently high. In such case, the model degenerates to a model with indivisible labor discussed in [Section 6.2](#), which is possible to explain some, but not all, of sharp retirement behavior.

This paper is not trying to understate the importance of the discrete labor choices approach or the model with fixed costs in explaining sharp retirement. The point is that these two models, separately or combined, are able to explain some, but not all, sharp retirements in the data. The goal of this paper is to provide an alternative explanation. I believe that sharp retirement is not solely driven by any one of the explanations listed above. It is likely that sharp retirement is a result of all these factors and others which are still unknown to us. In order to understand retirement behavior better, especially in the currently rapidly aging populations of many countries, I believe it is important to explore all possible driving forces of sharp retirement.

7. Discussion

7.1 Comparing with Fixed Costs

This section derives and compares the implications of the working habit model and the fixed costs model.

The fixed costs model directly implies that, in order to minimize fixed costs, workers want to cluster labor supply within a working period no matter whether one period is defined as a year, month, or week.³⁰ For instance, working 40 hours per week for 20 weeks is more attractive than working 20 hours per week for 40 weeks since the former half of the fixed costs than the latter.³¹

³⁰Some literature assumes a re-entry cost but usually it not big compared to fixed costs. For example [French and Jones \(2011\)](#) estimate fixed time costs of re-entering the labor market at 5% of a yearly full-time job, or 0.37 hours per working day in the year of re-entry, which is only about 10% of fixed costs. So reducing working weeks instead of weekly hours is still preferred, even when paying for re-entry costs. Of course this is assuming the wage is constant within the period.

³¹Strictly speaking, for same h_t assuming the directly disutility $-\alpha_h \frac{h_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$ is independent of the allocation of h_t within

That is, if labor supply needs to be reduced to correspond to decaying productivity in a model with fixed costs, an individual would reduce weeks worked per year instead of hours worked per week within the single period, which is defined as one year in most literature. However this is not well supported in HRS data.³² Figure 8-9 show that individuals in HRS data are more likely to reduce hours worked per week rather than weeks worked per year.³³ For individuals choosing smooth retirement, the hours worked per week decrease from around 45 to below 20, a more than 50% deduction, while the number of weeks worked per year decreases from 50 to 40, only 20% lower. Similar patterns are found for individuals coming back to employment after retirement, where the average hours worked per week are around 25 while the average weeks worked per year are around 40 weeks, regardless of retirement type.

On the other hand, more consistent with the data, in a working habit model the effect of adjacent complementarity induces the individual to smooth labor supply within the period whenever it is optimal to work. A simple example illustrates this result. Assume the optimal labor supply at period t is h_t . Within this period compare two extreme cases: the smoothing case is to work $h_t/50$ per week for 50 weeks consecutively and the non-smoothing case is to work $h_t/25$ per week for every other week. Assume there are two discrete levels of working habit, $S_H > S_L$, and working $h_t/50$ per week is enough to sustain working habit of S_H . S_H decays to S_L after one non-working week and it also takes one-working week to increase from S_L to S_H . Then, apparently, the smoothing case generates a utility which is $\alpha_{hs} (S_H - S_L) h_t$ higher than the non-smoothing case. For the non-smoothing case, the amount of utility lost is the cost paid during the working habit re-accumulation period. Of course, this is an extreme case to illustrate how the working habit model predicts one individual would smooth labor supply within one working period. If the non-smoothing case is to work $h_t/25$ per week for the first 25 weeks, then the utility lost compared to the smoothing case is $\alpha_{hs} (S_H - S_L) h_t/25$, which is smaller but still positive.

These costs are similar to the fixed costs of re-entering the labor market from unemployment or retirement used in the literature (Rust and Phelan (1997), Altug and Miller (1998), and French and Jones (2011)). Comparing two individuals: worker E has always been working and working habit level is S_H while worker U is unemployed for a certain period and his working habit level drops to $S_L < S_H$. Upon returning to employment, before his working habit level approaches S_H , worker U gets lower utility than worker E . If one period is required to increase working habit from S_L to S_H then the utility difference is $\alpha_{hs} (S_H - S_L) h$.

A direct implication of the above discussion is that, in addition to wage loss, unemployed workers also suffer from decreasing working habit during unemployment in this model. That said, an implication of this model is that when a reduction in total labor demand is necessary (during an economic downturn, for instance), cutting back labor supply from each worker while

the period.

³²PSID data have lots of missing on hours worked per week or weeks worked per year. The measure of working weeks the year before retirement is not very relevant due to retiring mid-year on PSID and CPS. Thus I only investigate the HRS data on this matter.

³³The implication of reducing weeks worked per year instead of hours worked per week should hold for all individuals in a model with fixed costs regardless of how they retire since they only differ on the magnitude of fixed costs.

keeping the same employment is better than cutting back employment while keeping same labor supply from each employed worker, in terms of generating higher total utility. This helps justify the work-sharing program currently provided by the Employment Development Department (EDD) of California³⁴ and implemented in Germany.

Some might argue it is possible that individuals might cluster labor supply at the daily level within each week by cutting back days worked instead of hours worked per day. Unfortunately, the three data sets used in this paper do not indicate daily labor supply, but I argue even if the individuals cluster labor supply at the daily level, it is unlikely due to the fixed costs of working. This involves the question of what fixed costs really are and how often they are paid. If fixed costs are paid each week, such as some type of psychological costs, then workers should cut back weeks worked, which is not supported by data. If fixed costs are paid each day, such as commuting costs, it is optimal for workers to cut back days worked. This assumes fixed costs are proportional to days worked, and thus the yearly level fixed costs are also proportional to the yearly labor supply. In such a set up, fixed costs are equivalent to a deduction in wages,³⁵ and the model is similar to the standard model in Section 6, which finds difficulty in justifying sharp retirement.³⁶

7.2 Labor Supply Elasticities

Besides rationalizing sharp retirement, the working habit model also helps explain the discrepancy between micro and macro elasticities of labor supply by generating a very small elasticity on the intensive margin and a relatively larger elasticity on the extensive margin. The large elasticity on the extensive margin comes from workers quitting cold turkey while the small elasticity on the intensive margin comes from that the return to labor supply from the reduction of future marginal disutility of working induces workers to work extensively, even at relatively low wages. The estimated results of labor supply elasticities at different ages are summarized in Table 5.

The labor supply elasticity at age t_1 responding to wage changes at age t_2 , $\epsilon_{t_2}^{t_1}$, is calculated in a perfect foresight model as follows: increase wages at age t_2 by 10% and then calculate the percentage change of labor supply at age t_1 , $\Delta\%h_{t_1}$, which is divided by 10% to get the elasticity. Three different elasticities of labor supply are estimated. The macro elasticity, $\epsilon_{t_2}^{t_1}$, corresponds to changes in the total number of hours worked for all individuals. The micro elasticity, $\epsilon_{i,t_2}^{t_1}$, corresponds to changes in the average number of hours worked, conditional on working,³⁷ while the

³⁴<http://www.edd.ca.gov>

³⁵If fixed costs per time unit is w_{fc} then the wage income is $(w - w_{fc})h$ with per time unit wage w and labor supply h .

³⁶However, if individuals do cluster labor supply at the daily level, the working habit model has difficulty explaining it. The working habit model also fails to explain why individuals work 8 hours a day for 5 weeks instead of 40/6 hours a day for 6 days. As discussed above the working habit model predicts one individual would smooth labor supply within the working period. This is a limitation of the working habit model. One possible fix is to include two types of working habit: one is current working habit, accumulated at a low frequency and displays adjacent complementarity, the other is an opposite type, accumulated at a high frequency and displays adjacent substitution and thus needs to be cleared periodically, for example weekly. This is outside of the scope of this paper and left for future research.

³⁷Due to the wealth effect and the adjacent complementarity, the calculated labor supply elasticity on the intensive

elasticity on the extensive margin, $\epsilon_{e,t_2}^{t_1}$, is calculated by the change in the labor force participation rates.

Blundell and MaCurdy (1999) summarize that the estimated micro elasticities of labor supply are quite small, ranging from 0 to 0.05 for married men in PSID data. This is consistent with findings in this paper. Table 5 shows that the micro elasticity is quite small, mostly ranging from .023 to .056 (with one exception of .121 at age 66). The model also generates relatively larger elasticities on the extensive margin, from 0.101 to 0.639. The macro elasticities range from 0.132 at age 56 to 0.680 at age 66.

The model provides an alternative explanation for the discrepancy between micro and macro elasticities of labor supply in addition to current approaches surveyed comprehensively in Keane and Rogerson (2011). It also helps explain the finding in Meyer (2002) that single mothers' labor supply responses to EITC are almost exclusively on the extensive margin. However the estimated macro elasticities are still smaller than those used in the business cycle literature.

Table 5 also reports how individuals respond to predictable wage changes in the future. The labor supply elasticities one period prior to wage changes are of the same magnitude as the ones when wage changes are realized. This is consistent with the finding in French (2004) that the labor supply response to mis-measured but predictable wage changes in the PSID data is small, for the reason that workers respond as soon as the changes are predicted rather than realized.

8. Counterfactuals

Policymakers are interested in how changes of social security policies will affect individual behavior. In this section I conduct three counterfactual experiments. The first experiment sets the Early Retirement Age (ERA) one period later which is equivalent to setting ERA from age 64 to 66. The second experiment eliminates the Social Security earnings test. In the third experiment I consider reducing the Social Security benefits by 20%. In all three counterfactual experiments I investigate how the overall labor supply, retirement ages and Social Security application ages are affected, as well as if responses are different for individuals choosing different retirement types—sharp or smooth. The results from the counterfactual experiments are summarized in Table 6.

The second column of Table 6 displays results from the experiment of shifting ERA from 62 to 64. All results are relative to the baseline simulation. After the policy change, on average each individual works an additional 1.57 months per lifetime, delays retirement by 1.80 months and delays Social Security application by 1.20 months. Individuals choosing sharp retirement respond more. They work 2.32 more months and delay retirement by 2.16 months, relative to 0.96 more months and 1.20 months from those choosing smooth retirement. This is due to the fact that individuals choosing sharp retirement tend to retire earlier than those choosing smooth retirement (Figure 16). Thus shifting the ERA affects individuals choosing sharp retirement more significantly.

margin—the micro elasticity—is smaller than the Frisch labor supply elasticity. Assuming certainty and an interior solution, the Frisch elasticity of labor supply is γ .

For similar reasons, eliminating the Social Security earnings test has larger effects on labor supply and retirement ages for individuals choosing sharp retirement, even though the effects are all small. In this experiment, workers choosing sharp retirement supply 0.58 more months per life-cycle, delay retirement by 0.57 months and apply Social Security benefits 2.40 months earlier. On Average eliminating Social Security earnings test encourages individuals apply Social Security benefits around 2.52 months earlier. [French \(2005\)](#) estimates a much larger effect from this experiment. He uses PSID covering 1968 – 1997 when all workers under age 70 are subject to the earnings test. I use HRS covering 1992 – 2008 when such earnings test has been removed for workers aged 65 or older.

In the third experiment I consider decreasing the generosity of the Social Security via reducing the benefits by 20%. It drives individuals work an additional 8.64 months per lifetime regardless of retirement type. This large response of labor supply is mainly caused by the income effect. Individuals work harder to accumulate more assets to offset benefits loss after retirement. Individuals choosing sharp retirement delay retirement 8.40 months. By contrast, individuals choosing smooth retirement delay retirement only by 1.20 months. For them most adjustment on the labor supply comes from the intensive margin. Most individuals choosing smooth retirement have relatively low degree of adjacent complementarity therefore higher labor supply elasticity on the intensive margin. Individuals delay Social Security benefits application by 2.04 months on average, with 2.76 or 1.56 months for those choosing sharp or smooth retirement.

Among these three counterfactual experiments, the first two only affect individuals who retire at certain ages while the third experiment affects everyone in general. As life expectancy increases and the labor force participation rates increase for elderly, the effect of increasing ERA could get smaller while universally reducing the generosity of Social Security benefits could become larger.

9. Conclusion

Understanding how individuals retire is as important as understanding why they retire. It is critical for questions ranging from assessing the impacts of tax policies to evaluating Social Security policies. In this paper, I first document how individuals retire from the labor market in three widely used data sets (CPS, PSID, and HRS). As widely acknowledged, the majority of retirement incidence is accompanied by an abrupt and discontinuous decline in hours worked. That is, most individuals directly retire from their full-time jobs without going through any period of part-time work. Such sharp retirement is hard—if not impossible—to explain with a standard labor supply model, where both preference and productivity change gradually over time. In order to rationalize why individuals do not smooth labor supply in the transition to retirement, some level of non-concavity either in production or preference is required. Models with discrete labor choices or fixed costs share the same spirit—a non-concave production function. This paper proposes and estimates a different approach, which essentially assumes a non-concave preference with habit persistence. I argue this is a reasonable assumption and well-supported by previous literature and various data. Using HRS data, I estimate a life-cycle labor supply model with working habit

in which sharp retirement can be explained by workers quitting cold turkey. The model produces reasonable parameter estimates with good model fits.

This model provides an alternative explanation to rationalize sharp retirement, which is supported better by data than existing models. In particular, I show that, on average individuals gradually reduce hours worked per week and keep weeks worked per year relatively constant. This is the opposite of what a model with fixed costs would predict, but is consistent with a model featuring adjacent complementarity, as is proposed in the paper.

The model enables me to estimate how individual labor supply and retirement behavior respond to changes in Social Security rules. Counterfactual experiments suggest that increasing the Early Retirement Age or eliminating Social Security earnings test has a moderate or little effect on labor supply or retirement ages, while universally reducing the generosity of Social Security benefits has much larger effect on the labor supply and retirement ages due to the income effect. In particular, reducing Social Security benefits by 20% makes individuals work an additional 8.6 months per lifetime. Individuals choosing sharp retirement respond mostly on the extensive margin by delaying retirement eight months, while individuals choosing smooth retirement respond mostly on the intensive margin by increasing yearly labor supply and delaying retirement only one month.

References

- Altug, Sumru and Robert A. Miller**, "The Effect of Work Experience on Female Wages and Labour Supply," *The Review of Economic Studies*, January 1998, 65 (1), 45–85.
- Ashenfelter, Orley and David Card**, "Did the Elimination of Mandatory Retirement Affect Faculty Retirement?," *The American Economic Review*, September 2002, 92 (4), 957–980.
- Becker, Gary S. and Kevin M. Murphy**, "A Theory of Rational Addiction," *The Journal of Political Economy*, 1988, 96 (4), 675–700.
- Blau, David and Tetyana Shvydko**, "Labor Market Rigidities and the Employment Behavior of Older Workers," *Industrial and Labor Relations Review*, April 2011, 64 (3), 464–484. working paper, the Ohio State University.
- Blau, David M.**, "Retirement and Consumption in a Life Cycle Model," *Journal of Labor Economics*, January 2008, 26 (1), 35–71.
- Blundell, Richard and Thomas MaCurdy**, "Labor Supply: A Review of Alternative Approaches," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3 1999, chapter 27.
- Bover, Olympia**, "Relaxing Intertemporal Separability: A Rational Habits Model of Labor Supply Estimated from Panel Data," *Journal of Labor Economics*, January 1991, 9 (1), 85–100.
- Casanova, Maria**, "Happy Together: A Structural Model of Couples' Joint Retirement Choices," November 2010. Working paper.
- Chetty, Raj**, "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, 80 (3), 969–1018.
- Citro, Constance Forbes and Robert T. Michael**, *Measuring Poverty: A New Approach*, Washington, DC: National Academy Press, 1995.
- Cogan, John F.**, "Fixed Costs and Labor Supply," *Econometrica*, July 1981, 49 (4), 945–963.
- DeNardi, Mariacristina**, "Wealth Inequality and Intergenerational Links," *The Review of Economic Studies*, July 2004, 71 (3), 743–768.
- , **Eric French, and John Bailey Jones**, "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, February 2010, 118 (1), 39–75.
- Donald, Stephen G. and Daniel S. Hamermesh**, "A Structural Model of the Fixed Time Costs of Market Work," *Economics Letters*, September 2009, 104 (3), 125–128.
- E., Jr. Ryder Harl and Geoffrey M. Heal**, "Optimal Growth with Intertemporally Dependent Preferences," *The Review of Economic Studies*, January 1973, 40 (1), 1–31.

- French, Eric**, "The Labor Supply Response to (Mismeasured but) Predictable Wage Changes," *Review of Economics and Statistics*, May 2004, 86 (2), 602–613.
- , "The Effects of Health, Wealth and Wages on Labor Supply and Retirement Behavior," *Review of Economic Studies*, April 2005, 72 (2), 395–427.
- **and John Bailey Jones**, "The effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, May 2011, 79 (3), 693–732.
- Gonzalez, Jorge**, "Commuting Costs and Labor Force Retirement," 2008. IVIE working paper WP-AD 2008-19.
- Gustman, Alan L. and Thomas L. Steinmeier**, "A Structural Retirement Model," *Econometrica*, May 1986, 54 (3), 555–584.
- **and** – , "Social Security and Retirement Dynamics," July 2006. Final Report, Michigan Retirement Research Center, UM05-05.
- **and** – , "How Changes in Social Security Affect Recent Retirement Trends," June 2008. NBER working paper 14105.
- Heckman, J. and B. Singer**, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, March 1984, 52 (2), 271–320.
- Hotz, V. Joseph, Finn E. Kydland, and Guilherme L. Sedlacek**, "Intertemporal Preferences and Labor Supply," *Econometrica*, March 1988, 56 (2), 335–360.
- Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes**, "Precautionary Saving and Social Insurance," *The Journal of Political Economy*, April 1995, 103 (2), 360–399.
- Hurd, Michael D.**, "The Effect of Labor Market Rigidities on the Labor Force Behavior of Older Workers," in David Wise, ed., *Advances in the Economics of Aging*, University of Chicago Press for the NBER, 1996.
- Imai, Susumu and Michael P. Keane**, "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, May 2004, 45 (2), 601–641.
- Keane, Michael P. and Kenneth I. Wolpin**, "The Career Decisions of Young Men," *Journal of Political Economy*, June 1997, 105 (3), 473–522.
- **and Richard Rogerson**, "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective," September 2011. NBER Working Paper No. 17430.
- Kennan, John**, "An Econometric Analysis of Fluctuations in Aggregate Labor Supply and Demand," *Econometrica*, March 1988, 56 (2), 317–333.

- Kopecky, Karen A. and Richard M.H. Suen**, "Finite state Markov-chain approximations to highly persistent processes," *Review of Economic Dynamics*, July 2010, 13 (3), 701–714.
- Kubin, Ingrid and Aloys Prinz**, "Labour supply with habit formation," *Economics Letters*, March 2002, 75 (1), 75–79.
- Lazear, Edward P.**, "Why Is There Mandatory Retirement?," *The Journal of Political Economy*, December 1979, 87 (6), 1261–1284.
- Meyer, Bruce D.**, "Association Labor Supply at the Extensive and Intensive Margins: The EITC, Welfare, and Hours Worked," *American Economic Review*, May 2002, 92 (2), 373–379.
- Miller, Robert A. and Seth G. Sanders**, "Human capital development and welfare participation," *Carnegie-Rochester Conference Series on Public Policy*, 1997, 46, 1–43.
- Prescott, Edward C., Richard Rogerson, and Johanna Wallenius**, "Lifetime aggregate labor supply with endogenous workweek length," *Review of Economic Dynamics*, 2009, 12, 23–36.
- Rogerson, Richard and Johanna Wallenius**, "Fixed Costs, Retirement and the Elasticity of Labor Supply," October 2010. mimeo, Arizona State University.
- Rust, John and Christopher Phelan**, "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica*, July 1997, 65 (4), 781–831.
- Scholz, John Karl, Ananth Seshadri, and Surachai Khittrakun**, "Are Americans Saving "Optimally" for Retirement?," *Journal of Political Economy*, 2006, 114 (4), 607–643.
- Woittiez, Isolde and Arie Kapteyn**, "Social interactions and habit formation in a model of female labour supply," *Journal of Public Economics*, November 1998, 70 (2), 185–205.

Figure 1: [CPS] Yearly (left) and weekly (right) labor supply at the year prior to retirement.

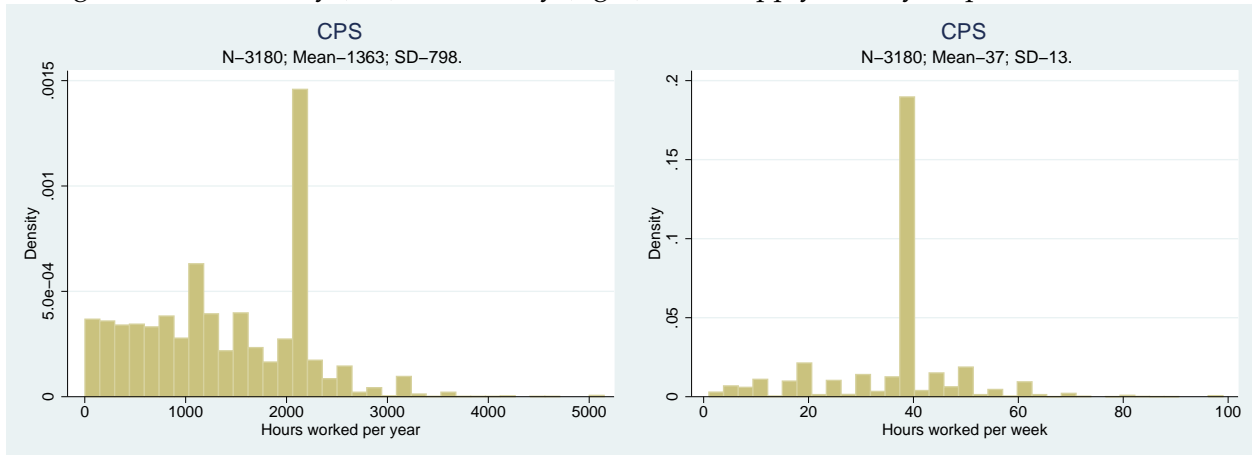


Figure 2: [CPS] Weekly labor supply profiles of workers choosing sharp or smooth retirement.

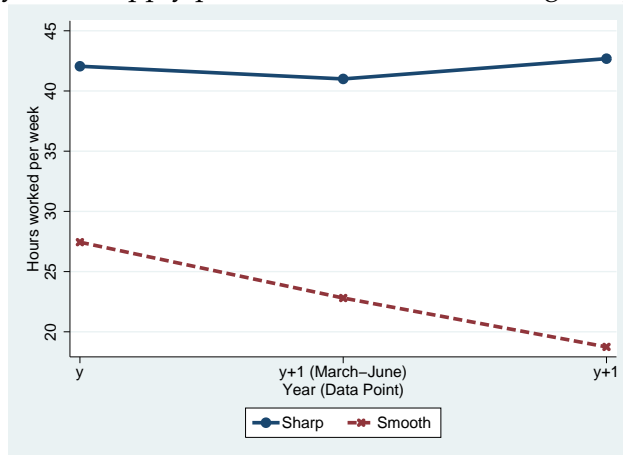


Figure 3: [CPS] Ratio of full-time working (left) and weekly hours (right) of workers aged 30-50 and working FT/PT at year $y + 1$ (March-June)

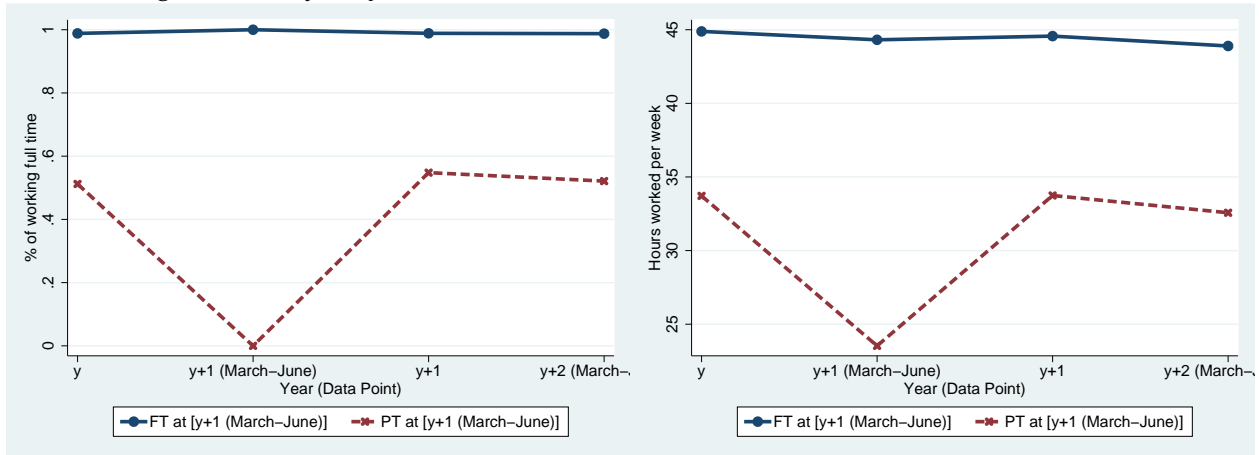


Figure 4: [CPS] Yearly (left) and weekly (right) labor supply of workers aged 30-50.

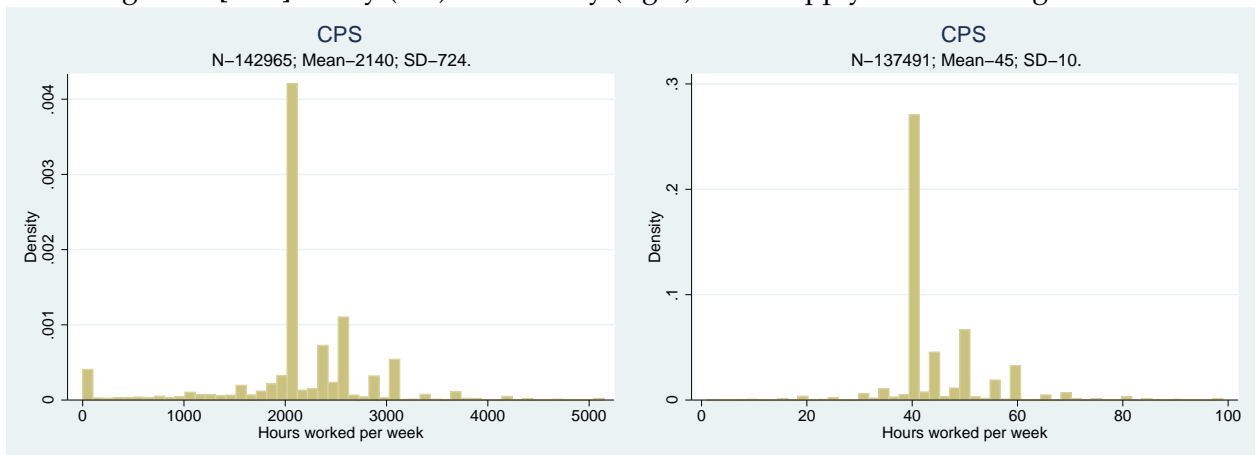


Figure 5: [PSID] Yearly labor supply, two years [left] or one year [right] before retirement.

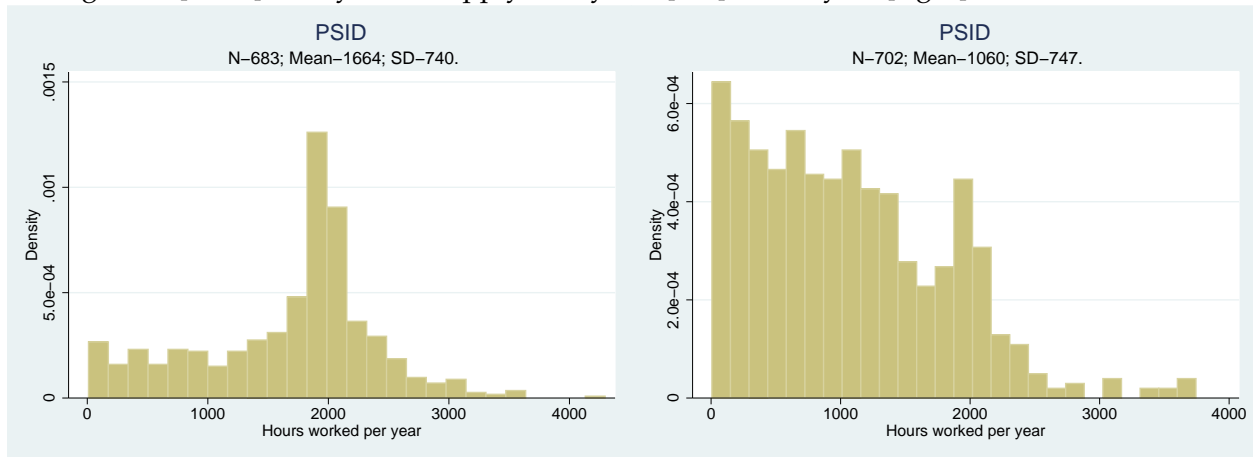


Figure 6: [PSID] Yearly labor supply.

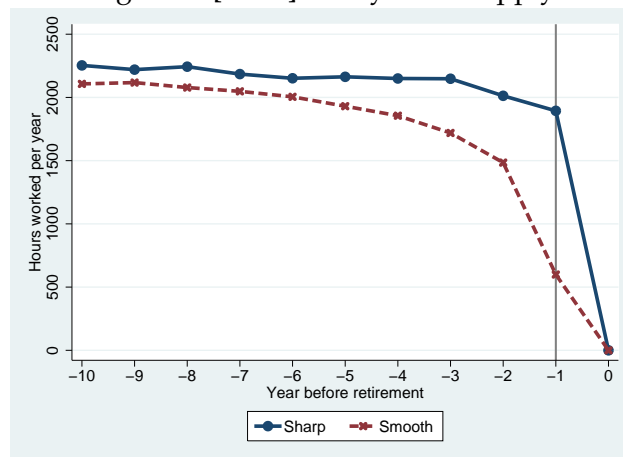


Figure 7: [HRS] Yearly labor supply, four years [left] or two years [right] before retirement.

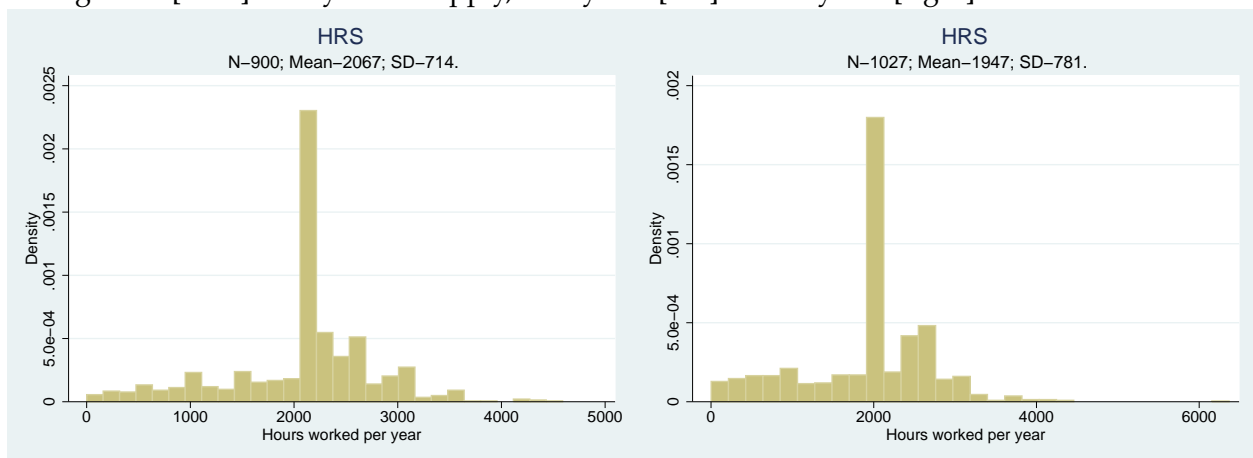


Figure 8: [HRS] Yearly labor supply over years relative to First Full Retirement Wave (FFRW).

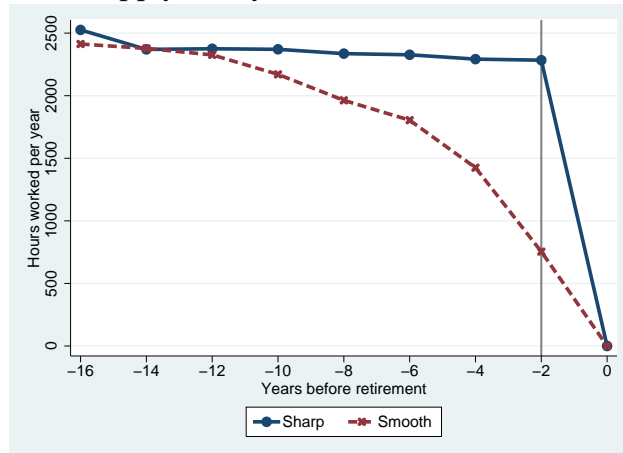


Figure 9: [HRS] Hours worked per week (left) and weeks worked per year (right).

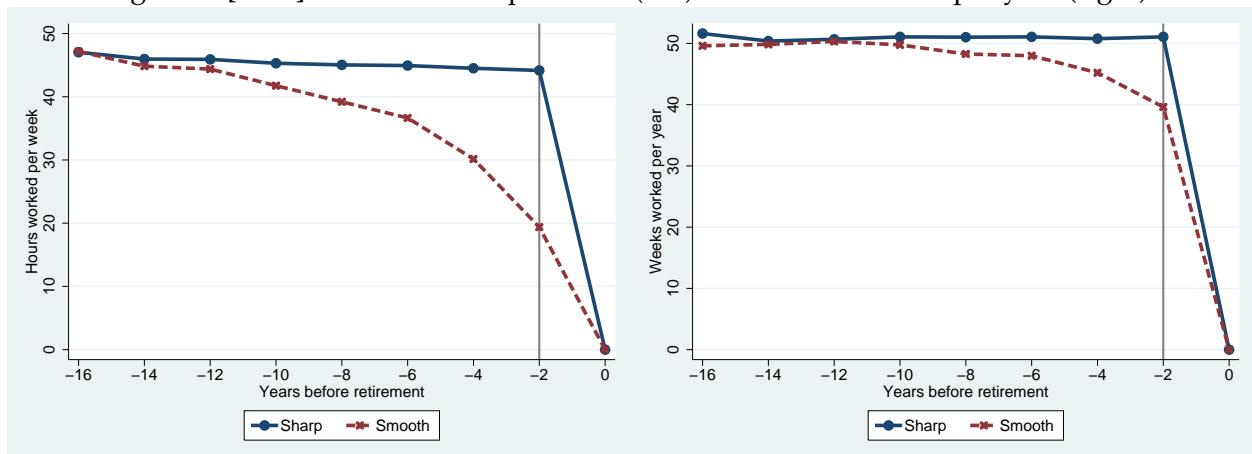


Table 2: Ratio of full-time working before retirement at CPS, PSID and HRS data.

%	CPS				PSID		HRS	
	Weekly (hours)		Yearly (hours)		Weekly	Yearly	Weekly	Yearly
	≥ 35	≥ 20	$\geq 35 \times 45$	$\geq 35 \times 36$	≥ 35	$\geq 35 \times 36$	≥ 35	$\geq 35 \times 36$
Hours								
Overall	80.2/75.1	91.1/89.1	70.9/42.5	75.4/52.5	88.9/82.1	74.9/35.8	83.0/76.9	86.6/80.7
Married	80.7/75.6	91.0/89.4	71.8/43.4	76.0/53.4	89.0/82.2	75.3/35.5	83.0/76.6	86.3/80.7
Single	76.7/71.8	91.8/87.3	65.4/37.2	71.3/46.4	88.5/81.8	70.4/38.9	83.2/78.8	88.0/80.7
Blue collar	82.7/78.4	93.0/91.3	71.0/41.3	75.7/51.2			83.2/78.5	87.6/83.1
White collar	80.6/75.5	90.1/88.2	73.6/46.0	77.7/56.1			82.7/77.3	85.6/79.6
Good health	80.1/75.2	91.3/90.2	72.9/43.7	76.5/53.9	94.7/90.5	88.2/47.6	82.8/77.8	86.2/80.7
Bad health	68.8/66.7	87.9/85.3	64.5/36.2	70.9/46.3	81.6/72.7	70.0/43.2	84.5/71.4	89.3/80.6
Not self-emp	83.3/79.2	92.6/91.2	74.5/44.4	78.3/54.5	91.7/86.8	76.1/36.4	85.9/80.8	88.6/84.2
Self-emp	58.9/47.0	81.3/74.7	47.0/30.3	55.7/38.6	74.7/60.4	66.7/31.8	61.2/49.1	71.6/54.6

Note: in each entry with r_{-2}/r_{-1} , r_{-2} is the ratio at two interview points before the retirement point, and r_{-1} is one interview point before.

Figure 10: An example with deterministic linear wage from 5 to 0, from period 30 to 80.
 $r = 0.03$, $\beta(1+r) = 1$, $\rho = 3.0$, $\gamma = 0.2$, $\alpha_h = 30.0$, $\alpha_{hs} = 5.0$, $\alpha_s = 3.0$, $\delta = 1.0$, $A_{30} = 0.0$, $S_{30} = 0.4$.

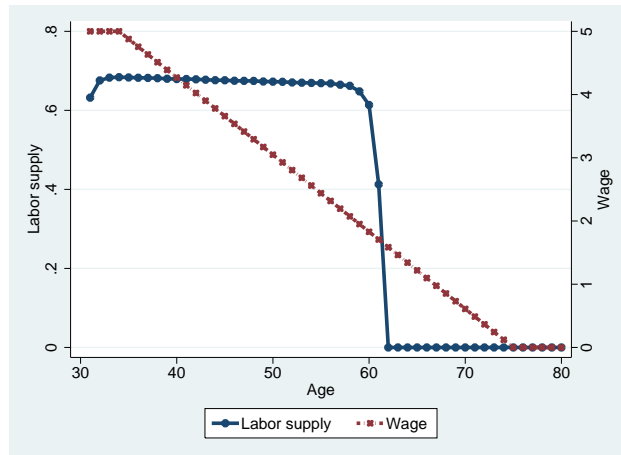


Table 3: Descriptive statistics of initial conditions.

		Overall	Sharp retirement	Smooth retirement	Not retired
N		1,506	708	254	544
Age	<i>Mean</i>	57.6	57.8	58.7	56.8
	<i>S.E.</i>	(3.45)	(3.49)	(3.44)	(3.21)
Asset	<i>Mean</i>	165,001	160,492	186,907	160,640
(\$)	<i>S.E.</i>	(185,683)	(167,730)	(186,344)	(206,112)
Wage	<i>Mean</i>	21.50	21.98	20.09	21.42
(\$/hour)	<i>S.E.</i>	(12.19)	(11.55)	(12.20)	(12.95)
Labor supply	<i>Mean</i>	2,230	2,269	1,792	2,384
(hours/year)	<i>S.E.</i>	(766)	(645)	(937)	(748)

Figure 11: Model Fit: Labor force participation rates

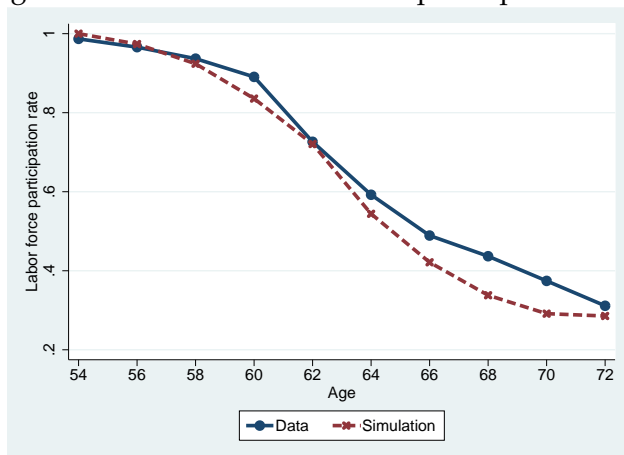


Figure 12: Model Fit: ratio of retirement type, sharp (left) or smooth (right)

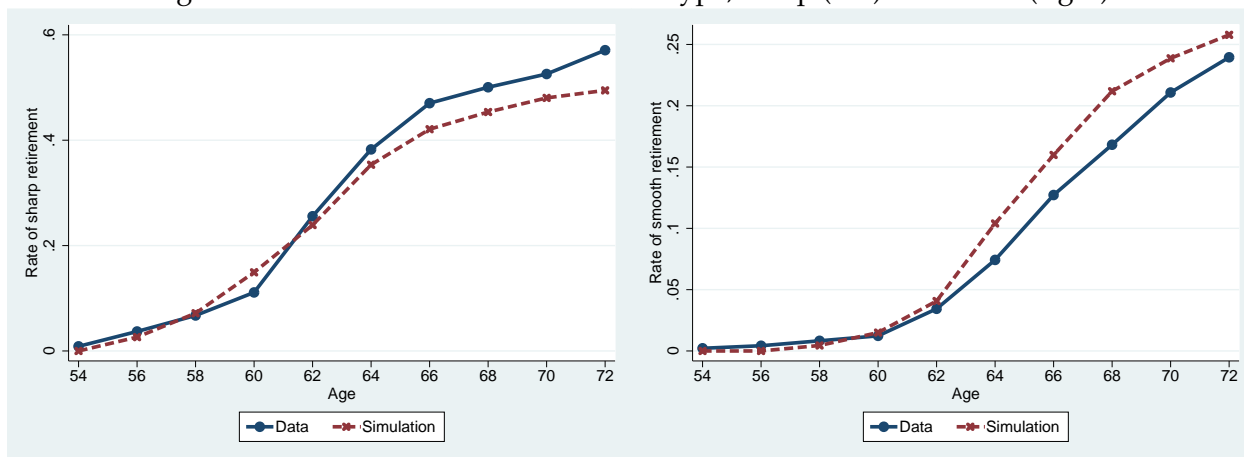


Figure 13: Model Fit: log wages, first (left) and second (right) moments

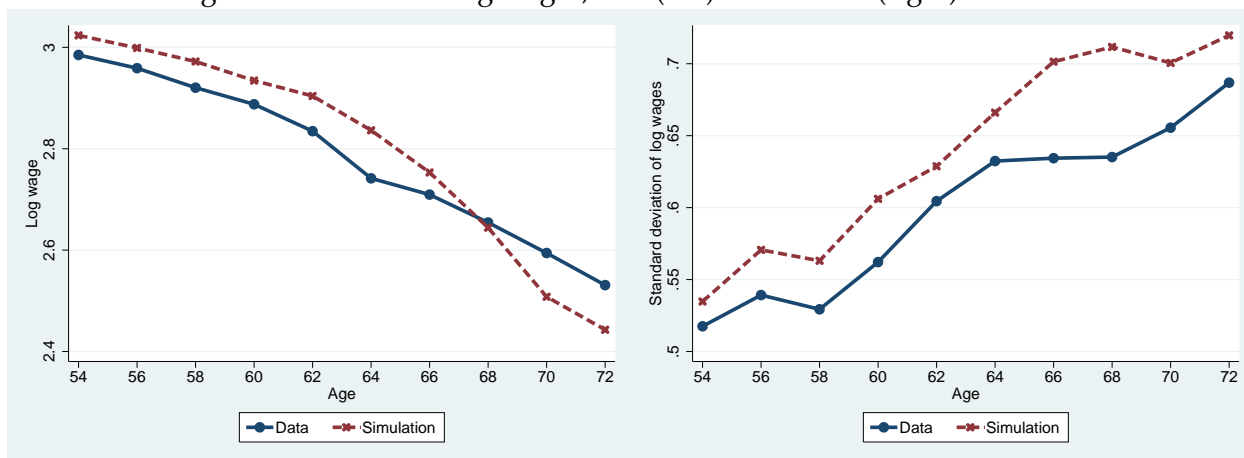


Figure 14: Model Fit: labor supply, first (left) and second (right) moments

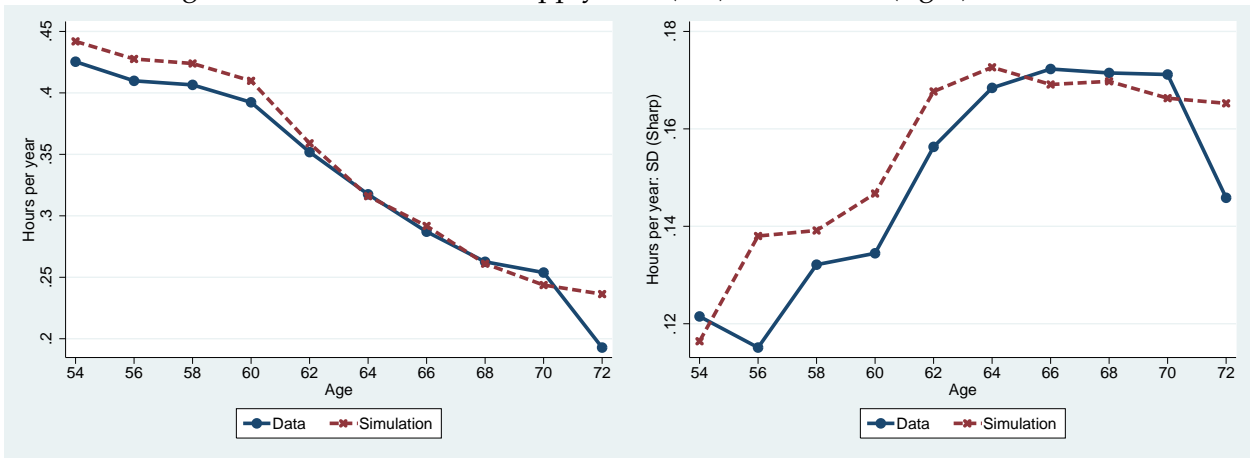


Figure 15: Model Fit: asset, first (left) and second (right) moments

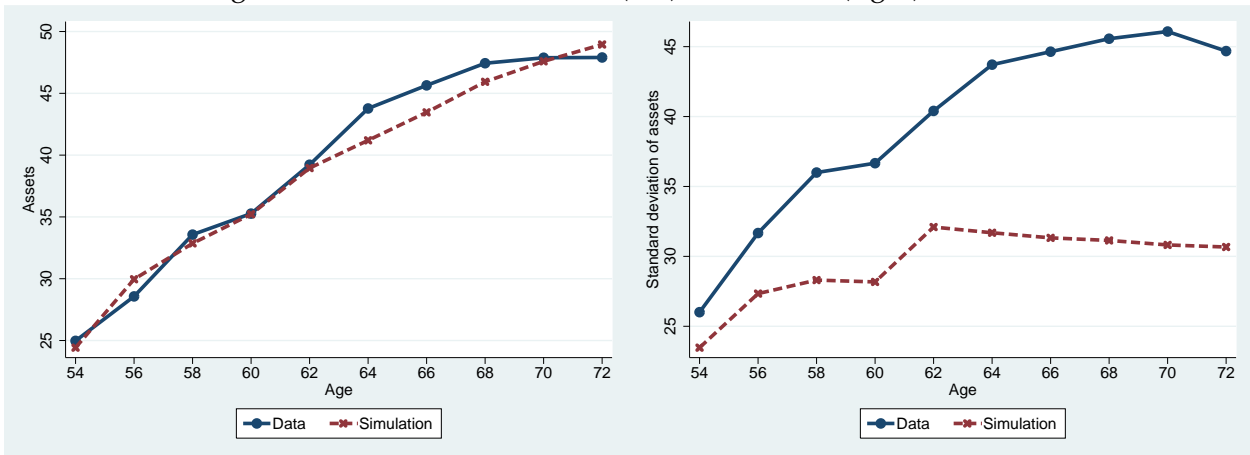


Figure 16: Age distributions of individuals choosing sharp (left) or smooth (right) retirement.

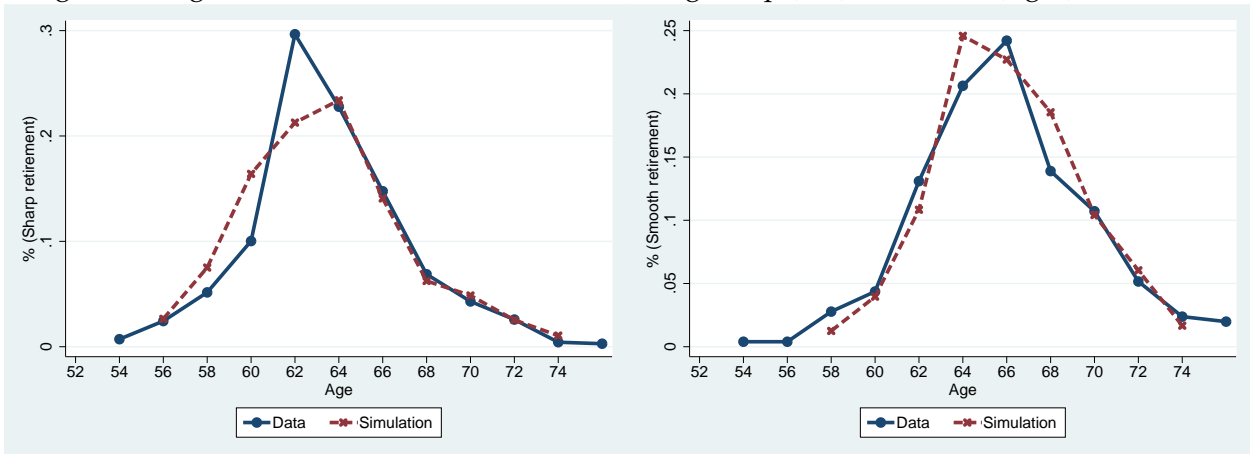


Figure 17: Labor supply of workers choosing sharp (left) or smooth (right) retirement.

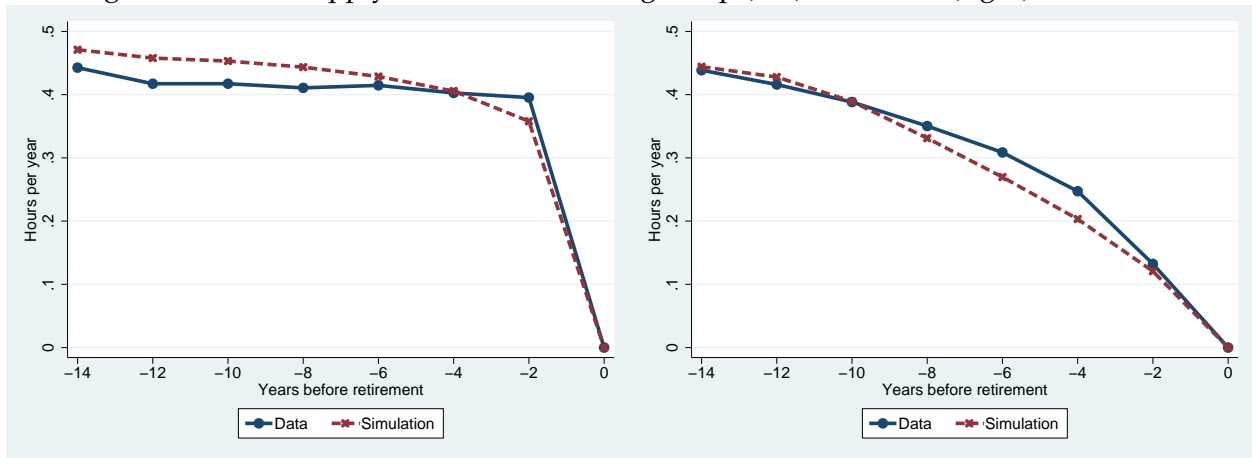


Figure 18: [CPS] Retirement age distribution before the year 2000 [Left] or after [Right].

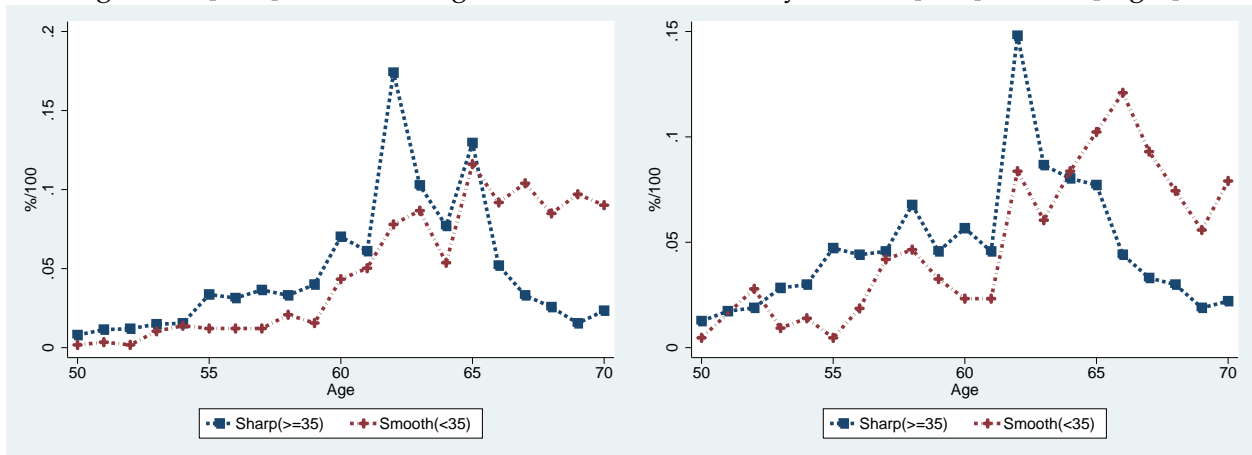


Figure 19: [HRS] Ratio of sharp retirement across occupations (left) or industries (right), on most recent jobs.

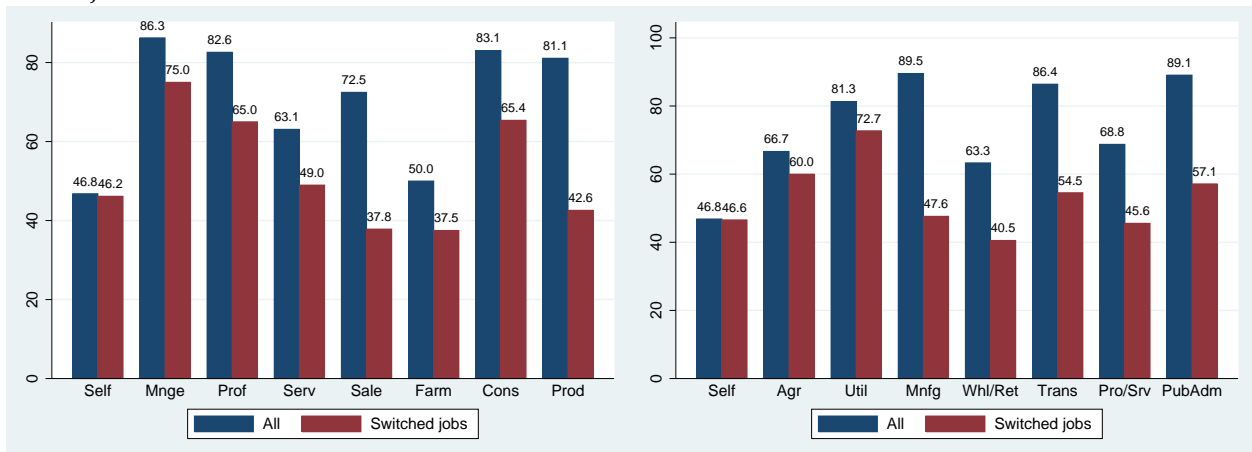


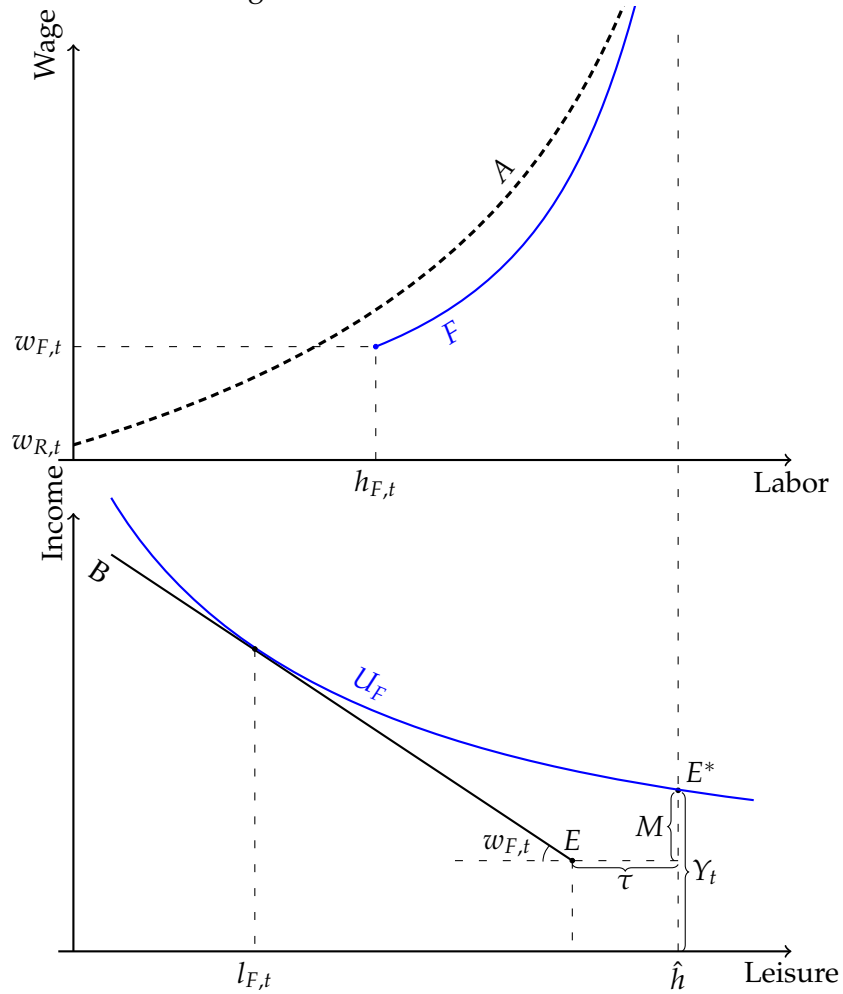
Table 4: Ratio of sharp retirement (%) on occupation switching.

Longest job	Most recent job						
	Self	Manag.	Prof.	Service	Sales/Office	Construction	Production
Management (%)	41.7	84.8	61.5	16.7	31.3	75.0	46.7
(#)	48	164	13	12	16	4	15
Professional	18.8	69.2	73.8	0.0	30.0	20.0	33.3
	32	13	145	2	10	5	6
Service	44.4	50.0	0.0	75.6	33.3		14.3
	9	2	1	41	3		7
Sales/Office	32.4	70.0	50.0	50.0	77.6	100.0	30.8
	34	10	2	10	125	3	13
Construction	36.1	100.0	66.7	46.7	0.0	86.8	45.5
	36	3	3	15	3	91	11
Production	42.4	100.0	33.3	44.0	57.1	75.0	87.3
	33	1	3	25	14	12	244

Table 5: Estimated labor supply elasticities based on 10% wage increase at age t .

Elasticity Age t	Macro		Micro (intensive margin)		Extensive margin	
	ϵ_t^{t-1}	ϵ_t^t	$\epsilon_{i,t}^{t-1}$	$\epsilon_{i,t}^t$	$\epsilon_{e,t}^{t-1}$	$\epsilon_{e,t}^t$
56	0.010	0.132	0.010	0.022	0.000	0.109
58	0.077	0.149	-0.020	0.047	0.097	0.101
60	0.072	0.231	0.025	0.023	0.046	0.207
62	0.115	0.461	0.078	0.056	0.037	0.404
64	0.181	0.663	0.090	0.024	0.090	0.639
66	0.269	0.680	0.026	0.121	0.242	0.558
68	0.160	0.571	-0.001	0.028	0.162	0.542

Figure 20: A model with fixed costs.



Note: Curve *A* - standard labor supply model;
 Curve *F* - model with fixed costs;
 Curve U_F - indifference curve in a model with fixed costs;

Table 6: Counterfactual experiments.

	Shift ERA from 62 to 64	Eliminate SS earnings test	Reduce SS by 20%
Labor supply per lifetime (months)			
Overall	1.57	0.28	8.64
Sharp	2.32	0.58	8.82
Smooth	0.96	0.10	8.43
Retirement age (months)			
Overall	1.80	0.24	5.88
Sharp	2.16	0.57	8.40
Smooth	1.20	0.09	1.20
SS application age (months)			
Overall	1.20	-2.52	2.04
Sharp	1.20	-2.40	2.76
Smooth	1.20	-2.76	1.56

SS stands for Social Security.

All entries are relative to simulation from the baseline model. For example, eliminating SS earnings test makes workers on average supply 0.28 more months than in the baseline model, but it also makes workers starting receiving SS benefits 2.52 months earlier (as the entry -2.52 indicates.)

Table 7: Merge MORG CPS with March CPS.

DP 1 (March CPS)		DP 2 (MORG CPS)		DP 3 (March CPS)		DP 4 (MORG CPS)	
year y		year $y + 1$ (March-June)		year $y + 1$		year $y + 2$ (March-June)	
mis	month	mis	month	mis	month	mis	month
1	<i>March</i>	4	<i>June</i>	5	<i>March</i>	8	<i>June</i>
2	<i>March</i>	4	<i>May</i>	6	<i>March</i>	8	<i>May</i>
3	<i>March</i>	4	<i>April</i>	7	<i>March</i>	8	<i>April</i>
4	<i>March</i>	4	<i>March</i>	8	<i>March</i>	8	<i>March</i>

Note: mis: Month in Survey, 1-8. DP: data point.

Table 8: Sample: White males 50-70 year-old at data point 4 in the merged data.

Interview Month in MORG	Before Year 2000	After Year 2000	Total
March	594	175	769
April	619	229	848
May	588	234	822
June	617	253	870

A. Data

All the nominal values of wages and assets are deflated to 2004 real values.

A.1 Current Population Survey (CPS)

The CPS merged outgoing rotation groups (MORG) data are merged with the CPS Annual Demographic File (March) data to get a short panel with four data points for each individual (Table 7). MORG CPS collects labor market participation information (labor force status, hours, hourly/weekly earnings) for last week while March CPS collects those information for last year. March CPS collects yearly earnings instead of hourly/weekly earnings. Both data include some demographic information. Some overlapping between DP3 and DP2 is possible since DP3 collects information for the prior year.

The sample is restricted to white males aged between 50 and 70 who are observed only retired at their fourth data point which is the last time they are observed in the CPS survey. Individuals with top 2% wages or more, or with bottom 2% wages or less at any point are excluded. Table 8 lists the sample size. There are totally 3,309 observations.

A.2 Panel Survey of Income Dynamics (PSID)

The Panel Survey of Income Dynamics (PSID) data is a longitudinal survey of U.S. families and their members, primarily collecting economic and demographic information with substantial detail on income, employment and family structure. From 1968 to 1997 individuals in PSID are

interviewed and reinterviewed every year. Since 1997, the survey is conducted biennially.

In this paper, data between 1968 and 1997 are selected when the survey is conducted yearly for a better observation of employment to retirement transition. The sample is restricted to white male household head only. Individuals with top 2% wage or more, or bottom 2% wage or less at any point are excluded. The final sample includes 106,830 observations for 8,770 individuals over 31 years.³⁸

A.3 Health and Retirement Survey (HRS)

The Health and Retirement Study (HRS) is a national panel survey of individuals over age 50 and their spouses, elicits information about demographics, income, assets, job status and history, family structure and many others. This paper uses the initial HRS cohort only who were born 1931 to 1941. This cohort was first interviewed in 1992 at age 51 to 61 and subsequently every two years. The latest interview wave selected for the paper is wave 9 in 2008. This paper primarily uses RAND HRS data, version K.³⁹

The following sample selection rules are applied in order:

- Exclude individuals with top 2% asset/wage or more, or with bottom 2% asset/wage or less at any point.
- White non-Hispanic male only.
- Exclude observations ever applied⁴⁰ or receiving Supplemental Security Income (SSI) or Social Security disability SSDI, or ever disabled in any wave.
- Exclude individuals who receive Social Security benefits on or before age 61.

Missing values on assets, Social Security income benefits, AIME and wages (for workers) are imputed as in [French and Jones \(2011\)](#).

B. Taxes

Worker's wage income is subject to the federal and state income taxes, and the payroll taxes.

The payroll taxes include the Social Security portion, 6.2% up to a limit which varies each year. I use the limit of 2004 which is \$87,900. The Medicare tax which is 1.45%, uncapped.

The federal and state income taxes are progressive. I ignore the state income taxes and only model the federal income tax. I assume everyone is subject to 2004 federal income tax rules under head of household.

³⁸Not all individuals enter the sample from 1968.

³⁹"RAND HRS Data, Version K. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (March 2011)." "The RAND HRS Data file is an easy to use longitudinal data set based on the HRS data. It was developed at RAND with funding from the National Institute on Aging and the Social Security Administration."

⁴⁰I expect that workers who applied SSI or SSDI might be systematically different from other groups of workers.

Table 9: Income Tax Codes.

Marginal Tax Rate	Pre-tax (Y)	Post-tax Income
0.0765	$\leq 10,250$	$0.9235Y$
0.1765	$10,251 - 20,450$	$9,465.88 + 0.8235(Y - 10,250)$
0.2265	$20,451 - 49,150$	$17,865.58 + 0.7735(Y - 20,450)$
0.3265	$49,151 - 87,900$	$40,065.03 + 0.6735(Y - 49,150)$
0.2645	$87,901 - 110,750$	$66,163.15 + 0.7355(Y - 87,900)$
0.2945	$110,751 - 172,950$	$82,969.33 + 0.7055(Y - 110,750)$
0.3445	$172,951 - 329,350$	$126,851.43 + 0.6555(Y - 172,950)$
0.3645	$\geq 329,351$	$229,371.63 + 0.6355(Y - 329,350)$

Table 10: Normal Retirement Age (NRA).

Year of birth	NRA	Year of birth	NRA
≤ 1937	65	1955	66 and 2 months
1938	65 and 2 months	1956	66 and 4 months
1939	65 and 4 months	1957	66 and 6 months
1940	65 and 6 months	1958	66 and 8 months
1941	65 and 8 months	1959	66 and 10 months
1942	65 and 10 months	≥ 1960	67
1943-54	66		

The personal exemption for each person is \$3,100. The standard deductions for head of household is \$7,150.

These two taxes together generate the following tax code used in the paper in Table 9.

C. Social Security

Most of information about Social Security benefits in this section are from <http://www.ssa.gov>.

C.1 Normal Retirement

The worker receives full Social Security benefits if he begins receiving benefits at the normal retirement age (NRA). The NRA varies by year of birth as listed in Table 10.

The basic Social Security benefits is called the Primary Insurance Amount (PIA). At the NRA, the worker receives retirement benefits equal to the PIA. PIA is a function of Average Indexed Monthly Earnings (AIME),

$$PIA = 0.9 * \min \{bp_1, AIME\} + 0.32 * \min \{bp_2 - bp_1, \max \{0, AIME - bp_1\}\} + 0.15 * \max \{0, AIME - bp_2\} \quad (18)$$

Table 11: Bend points in PIA formula, nominal.

Year	bp_1 (\$)	bp_2 (\$)	Year	bp_1 (\$)	bp_2 (\$)
1990	356	2145	2000	531	3202
1991	370	2230	2001	561	3381
1992	387	2333	2002	592	3567
1993	401	2420	2003	606	3653
1994	422	2545	2004	612	3689
1995	426	2567	2005	627	3779
1996	437	2635	2006	656	3955
1997	455	2741	2007	680	4100
1998	477	2875	2008	711	4288
1999	505	3043			

The bend points bp_1, bp_2 vary by year of entitlement as listed in Table 11. For simplicity, I use the 2004 bend points $(bp_1, bp_2) = (612, 3689)$ for all the individuals in my sample.

Only Social Security benefits is observed in the HRS data. According to the mapping between PIA and AIME I calculate the AIME from the PIA

$$\begin{aligned}
 AIME &= \min \left\{ bp_1, \frac{PIA}{0.9} \right\} + \min \left\{ bp_2 - bp_1, \max \left\{ 0, \frac{PIA - 0.9 * bp_1}{0.32} \right\} \right\} \\
 &\quad + \max \left\{ 0, \frac{PIA - (0.9 * bp_1 + 0.32 * (bp_2 - bp_1))}{0.15} \right\} \\
 &= \frac{\min \{550.8, PIA\}}{0.9} + \frac{\min \{984.64, \max \{0, PIA - 550.8\}\}}{0.32} \\
 &\quad + \frac{\max \{0, PIA - 1535.44\}}{0.15}
 \end{aligned} \tag{19}$$

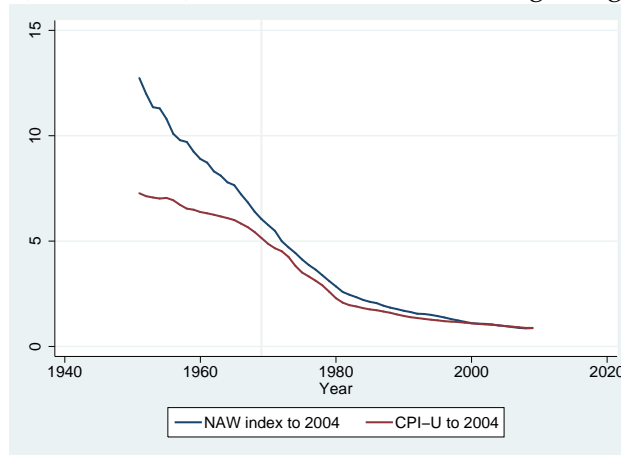
AIME is computed as the monthly average earning of the 35 years with highest inflation-adjusted earnings. Only earnings subject to the Social Security tax are used in the calculation therefore AIME is capped (Section B). The included earning in a specific year is adjusted for inflation by multiplying by the inflation or growth rate relative to the base year which is two years prior to the year of first eligibility - 62 for retirement. The inflation rate is calculated by dividing the average wage in the base year by the average wage in that specific year. Earnings after the base year (60 for retirement) are not adjusted.

The growth rate of the national average wage index is very similar to the growth rate of CPI-U after Year 1969 (Figure 21), so I ignore the small difference between these two and use the real wages to update AIME without adjustment.

Computing exact AIME requires keeping tracking of the worker's earning history. Without having the worker's entire earning history we have to apply an approximating method taking into account the wage growth pattern over life-cycle

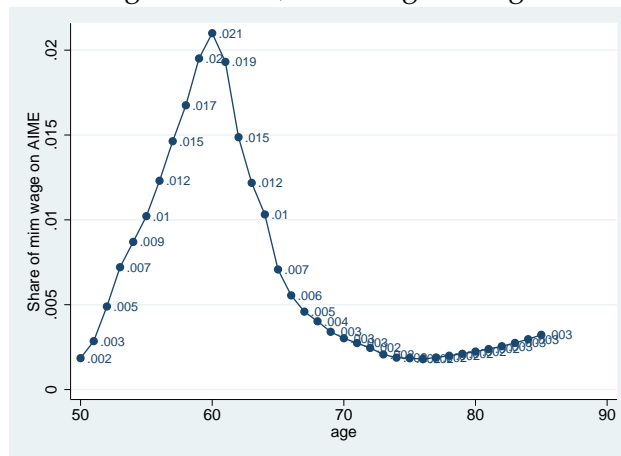
$$AIME_{t+1} = f_t^a (AIME_t, I_t, L_t) = AIME_t + \max \{0, sse_t/35 - 1 \{tw \geq 35\} \cdot share_{min} \cdot AIME_t\} \tag{20}$$

Figure 21: Relative (to Year 2004) indices of National Average Wage Index and CPI-U.



where $sse_t = \min \{wh_t, s\bar{s}e\}$ is included earning and $s\bar{s}e$ is the cap amount (\$87,900 on 2004). The tw is the number of years the worker has been working and $share_{min}$ is the share of minimum wage in AIME. Figure 22 lists the estimated $share_{min}$ from CPS data for age 52 to 76, assuming the starting working age of 16. All numbers are in 2004 dollar.

Figure 22: Share of minimum wage on AIME, assuming starting working from age 16. CPS data.



C.2 Early Retirement

The worker can also start to receive the Social Security benefits as early as age 62 (early retirement age or ERA) at a reduced level.

For early retirement, the benefit is reduced 5/9 of one percent for each month before NRA, up to 36 months. Beyond 36 months, the benefit is further reduced 5/12 of one percent per month.

Table 12: Early Retirement Benefit.

ERA relative to NRA (Year)	SS Benefits relative to Full Benefit
-1	(100-6.67)%
-2	(100-13.33)%
-3	(100-20)%
-4	(100-25)%
-5	(100-30)%

Table 13: Delayed Retirement Credit (DRC).

Year of birth	DRC	Year of birth	DRC
1917-24	3.0%	1935-36	6.0%
1925-26	3.5%	1937-38	6.5%
1927-28	4.0%	1939-40	7.0%
1929-30	4.5%	1941-42	7.5%
1931-32	5.0%	≥1943	8.0%
1933-34	5.5%		

C.3 Delayed Retirement

Delayed retirement after the NRA increases benefits. The delayed retirement credit (DRC) listed in Table 13 is given to the retiree for each delayed year up to age 69. No DRC is given for retirement at age 70 or older. For example, if a worker born on June 1935 retired on June 2002 then the benefit is $100\% + (67 - 65) \times 6.0\% = 112\%$ of her/his PIA.

For simplicity, I assume workers are all subject to DRC rate of 6.0%.

C.4 Earnings Test

The Social Security benefits could be withheld partly or totally if one worker is earning income while taking the Social Security benefits at ages before 70. The earnings test exempt amounts are listed in Table 14. Before 2000, \$1 of benefits for every \$2 of earnings in excess of the exempt amount is withheld for beneficiary under age 65, up to the total amount. The benefit withholding rate for those aged 65-69 is \$1 of benefits for every \$3 of earnings in excess of the exempt amount. Since 2000, the earnings test is eliminated after reaching NRA. Before reaching NRA the benefit withholding rate is \$1 benefit for every \$2 earnings in excess of the lower amount and at the year reaching NRA the “higher amount” every \$3 of earnings in excess of the higher amount is subject to being withheld \$1 in benefits applying to months prior to NRA.

If a whole year’s worth of benefits is withheld between ages 62 to 65, benefits in the future will be raised by 6.7% each year. If the benefit is withheld between age 66 to 70, the future benefits will be raised by the same amount shown in Table 13.⁴¹

⁴¹I would like to thank Maria Casanova for the help in understanding earnings test on this part.

Table 14: Annual Retirement Earnings Test Exempt Amounts.

Year	Under age 65	Age 65-69	Year	Lower amount	Higher amount
1990	6,840	9,360	2000	10,080	17,000
1991	7,080	9,720	2001	10,680	25,000
1991	7,440	10,200	2002	11,280	30,000
1993	7,680	10,560	2003	11,520	30,720
1994	8,040	11,160	2004	11,640	31,080
1995	8,160	11,280	2005	12,000	31,800
1996	8,280	12,500	2006	12,480	33,240
1997	8,640	13,500	2007	12,960	34,440
1998	9,120	14,500	2008	13,560	36,120
1999	9,600	15,500			

Note: Before 2000, the benefit withholding rate is \$1 for every \$2 (under age 65) or \$3 (age 65-69). Since 2000, “lower amount” only applies to the ages before NRA and “higher amount” is applied to the months prior to NRA at the year workers reach NRA, and there is no earnings test for ages after NRA.

For simplicity, I assume workers are all subject to the year 2004 rule. That is, the lower exempt amount is \$11,640 before reaching NRA. There is no higher exempt amount of \$31,080 since the model is estimated at the yearly level. All amounts are in 2004 dollar value. I ignore the earnings test for workers aged 65 or older before the year of 2000, which likely leads to overestimating the effect of working habit. The bias will not be big though. In HRS data, 43.5% of individuals retire before the year of 2000, of which only 15.0% are aged 65 or older with 64.7% choosing sharp retirement.

D. Other factors

D.1 Health Insurance

Health insurance is one of important facts that workers take into account when they decide the retirement timing, as studied in [Rust and Phelan \(1997\)](#) and [French and Jones \(2011\)](#). However, since I am focusing on how individuals retire rather than why they retire, I decide not to include the health insurances at this step. This compromise to the tractability will certainly include some bias in the estimation, mostly in the timing of retirement.

D.2 Mandatory Retirement

During the time covered in HRS data the mandatory retirement has been prohibited for most private sector jobs.

Lazear (1979) argues that the mandatory retirement results from an optimal wage profile which discourages shirking. In such model at the age of mandatory retirement the wage exceeds the worker's marginal productivity so it is optimal for the firm not to renew the contract at the current wage level, and the worker is better off to retire inasmuch as the reservation wage is higher than the marginal productivity. The latter assumption is questionable and if relaxed the worker does not necessarily have to retire. So mandatory retirement does not explain sharp retirement well either.

Ashenfelter and Card (2002) conduct a survey to estimate the effect of the elimination of mandatory retirement on faculties retirement behavior. They find that the retirement rates of faculties before age 70 are not affected by the elimination of mandatory retirement and that the retirement rates at age 70 and 71 drop more than half after 1994.