ARC Centre of Excellence in Population Ageing Research

Working Paper 2015/20

Retiring Cold Turkey

Xiaodong Fan[1]

[1]Research Fellow, Centre of Excellence in Population Ageing Research (CEPAR), Australian School of Business, University of New South Wales email: x.fan@unsw.edu.au

Retiring Cold Turkey

Xiaodong Fan*

January 8, 2015

Abstract

This article documents "sharp retirement" among white male workers in the United States—retirement accompanied by a discontinuous decline in labor supply. It then proposes and estimates a life-cycle model with habit persistence to explain such precipitous decline in labor supply upon retirement as workers quitting "cold turkey" to break the "work habit." Counterfactuals reveal heterogeneous responses from different retirement types. In response to reducing Social Security benefits by 20%, individuals choosing sharp retirement respond mostly on the extensive margin by delaying retirement eight months, while individuals retire smoothly respond mostly on the intensive margin by increasing yearly labor supply and delaying retirement only one month. Comparison shows the work habit model produces more empirically plausible results than other approaches.

KEYWORDS: retirement behavior, work habit, habit persistence, cold turkey

JEL Classification: J22, J26, J32

*ARC Centre of Excellence in Population Ageing Research (CEPAR), University of New South Wales (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, Christopher Taber and Alan Woodland for their support throughout the development of this article. I am grateful to Maria Casanova, Steven Durlauf, Hanming Fang, Michael Hurd, Kenneth Judd, Michael Keane, Sang Yoon Lee, Rasmus Lentz, Vladimir Matveenko, Tom Mroz, Salvador Navarro, Cain Polidano, John Rust, Carl Sanders, and seminar and conference participants at Aalto, ANU, Auckland, Clemson, CMES, EIEF, Hanqing, Mannheim, Saskatchewan, Tokyo, UNSW, Wisconsin, and Wustl for helpful comments. I gratefully acknowledge financial support from the Robock award at the University of Wisconsin, and the ARC Centre of Excellence in Population Ageing Research (CEPAR). All errors are solely mine.

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 studies which use fixed costs to explain this phenomenon, I propose and estimate a work habit model which not only rationalizes such retirement behavior, but also delivers more empirically plausible results than the fixed costs explanation. The work habit model incorporates fixed time costs, and the estimate is much smaller than those in previous literature but comparable to actual commute time in the data.[1]

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,[2] while others 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.[3] This leads to a gradually declining labor supply-age 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 ap-

[1] The concepts of *work habit* and *fixed costs* are explained in details in Subsection 4.1 and Subsection 3.3.

[2] They work either full-time or part-time when first observed in the data.

[3] In this article only one cohort is investigated so *time*, *age* or *period* are used interchangeably.

proaches are widely used in the literature: discrete labor choices or fixed costs of working. Discrete labor choice models assume that individuals can only supply discrete amount of labor—full time, part time, or none. This approach could be seen as a reduced form implication of the fixed costs approach, which assumes that individuals have to pay fixed costs, either money or time, to work.[4] It generates an incentive to cluster labor supply. Empirical evidence of such fixed costs include commute time. For example in the PSID data on average a full-time worker pays 0.64 hours per day to commute between home and work. However, this is far from enough to explain sharp retirement. French (2005), French and Jones (2011), Rogerson and Wallenius (2013) 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. In this model, each worker accumulates work habit at work. The work habit reduces the marginal disutility of working, which makes working less uncomfortable. That is, the labor supplies of adjacent periods appear to be complement to one another. The presence of such adjacent complementarity has two effects on labor supply. On one hand, the optimal labor supply is not only determined by the current wage offer, but also is affected by past and future labor supply through accumulated work habit. This makes labor supply less sensitive to wage changes and therefore reduces the elasticity of labor supply on the intensive margin. This also helps explain the highly clustered hours worked in the data. On the other hand, the inclusion of adjacent complementarity makes the utility function not jointly concave in labor supply and work habit. As a result, interior solutions are not always optimal and the corner solution of zero labor supply becomes relevant. In response to expected retirement, the adjacent complementarity induces a fast and steep drop in labor supply, similarly as a "chain reaction." When the non-concavity and the chain reaction together

[4]This article 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.

generate a quick and discontinuous drop in labor supply upon retirement, it becomes optimal to stop working abruptly from a relatively high level of hours worked as wages decline. That is, retiring “cold turkey” appears to be the optimal way to break work habit. In other words, while one worker is working he is in a “working mode” and supplies a relatively high level of labor to take advantage of the adjacent complementarity.[5] If he plans to quit working and retire, he wants to get out of the working mode as quickly as possible.

This type 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 that a bang-bang solution is optimal. Their paper did not focus on the dynamics that could generate cold turkey quitting, while these dynamics are the main ingredients of the current paper.

The idea of adjacent complementarity in labor supply is also closely related to the literature of intertemporally nonseparable preferences in 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 (Hotz et al., 1988; Bover, 1991; Altug and Miller, 1998; Woittiez and Kapteyn, 1998; Kubin and Prinz, 2002). All the previous literature listed studies intertemporal nonseparability on the intensive margin, while this article applies it to the extensive margin to explain sharp retirement.

I use the method of simulated moments to match life-cycle profiles estimated with the HRS data. Heterogeneous preferences in the degree of adjacent complementarity are incorporated to improve the model fit. I match life-cycle profiles of assets, labor supply, labor force participation and reentry rates, wages, and frequencies of sharp and smooth retirements on

[5]This article 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 article. Interestingly, the retirement pattern of female workers is very similar as male workers.

the population level, as well as conditional moments for those 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. The fixed time costs are also incorporated into the work habit model. The estimated fixed time costs are 157 hours per year, which is comparable to the average commute time in the data (around 160 hours per year), but is much smaller than other estimates in previous literature (mostly over $1,000$ hours per year).

Understanding how individuals retire, or why individuals retire sharply, is important in two aspects. First, it helps us better understand individual behavior. Second, it is critical in conducting counterfactual policy experiments. For instance, when estimating the effects of policy changes in Social Security 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 in the Social Security benefits. Overall, reducing Social Security benefits by 20% makes an average individual work an additional 8.6 months over the lifetime, 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 both policy changes have moderate or little effect on labor supply or retirement ages.

I also present evidence showing that the data appear more favorable to the work habit approach than the fixed costs approach. The fixed costs model predicts that workers would cluster labor supply whenever possible to minimize fixed costs. Assume one period is defined as one year and fixed costs are paid each week at work. If an individual wants to reduce his labor supply for one period, then he would try to reduce weeks worked per year instead of hours worked per week.[6] In contrast, the work habit model implies that as long as the worker works he wants to smooth labor supply whenever possible to maintain a certain level of work habit to enjoy the adjacent complementarity. Note that this is different from the cold turkey quitting discussed earlier. In the work habit model, when one retires, he most likely quits cold turkey if the level of adjacent complementarity is strong, otherwise he quits smoothly. However, regardless of the heterogeneity in the level of adjacent complementarity, when one works, he wants to smooth the work habit through a continuous spell of working periods. 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 is more consistent with the work habit model than the fixed costs model.

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 reviews previous approaches. Section 4 develops the life-cycle labor supply model with work habit. Section 5 describes the data and estimation strategies. Section 6 presents estimation results. Section 7 conducts three counterfactual experiments. Section 8 discusses other implications of the model, and Section 9 concludes. Details of data, as well as

[6]If fixed costs are paid each day, this is equivalent to that the yearly fixed costs are proportional to yearly labor supply. In such a setup fixed costs are equivalent to a "tax" in wage, and then the model is similar to the standard neoclassical model which finds difficulty in justifying sharp retirement.

technical details, are all in the appendices.

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 Demographic File" (March) to get a short panel with four data points covering three years for each individual. 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. For further details regarding data construction, please refer to Appendix A.

One of the main findings is that across all three data sets, most white male workers are never observed working part-time before retiring. Over 80% 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. The average hours worked per year 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 $1,260$ hours or more, which is defined as full-time, in the year prior to their complete retirement.[7] If the full-time is defined as $35 \times 45 = 1,575$ hours or more a year, then the ratio is slightly smaller, 42.5%. The ratios

[7] Rand 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.

barely vary across marital status, work type, or health status (Table 1).

Figure 1: [CPS] Yearly [left] and weekly [right] hours, the year prior to retirement.

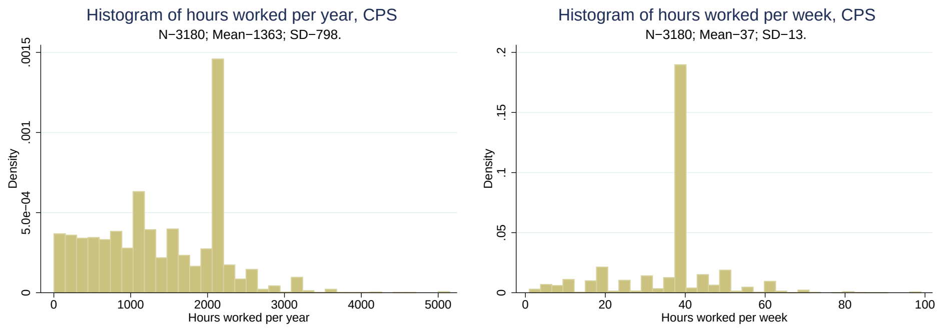


Table 1: Ratio of full-time working before retirement at CPS, PSID and HRS data.

	CPS				PSID		HRS	
%	Weekly (hours)		Yearly (hours)		Weekly	Yearly	Weekly	Yearly
Hours	≥ 35	≥ 20	$\geq 35 \times 45$	$\geq 35 \times 36$	≥ 35	$\geq 35 \times 36$	≥ 35	$\geq 35 \times 36$
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.

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, the left panel in Figure 2 plots the labor supply profiles for individuals who choose sharp versus smooth retirement in the CPS data. 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 (aged 30-50), where such a reduction of labor supply appears to be random. More than half of those workers working part-time at the second observed point in the CPS data are working full-time in the observations immediately before or after. For prime-age workers, 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.[8] At the yearly level, 35.8% of workers continue to work full-time the year before they completely retire from the labor market, which likely underestimates the prevalence of sharp retirement due to the mid-year retirement.[9] Indeed at the weekly level, the ratio of full-time workers goes up to 82.1% (Table 1). The dynamics of labor supply reveal that individuals choosing smooth retirement gradually cut back their hours worked per year before full retirement (Figure 2).

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. The

[8] Same as in the CPS data, in both the PSID data and the HRS data, at the yearly level "full-time" is defined as working $1,260$ hours or more per year and at the weekly level "full-time" is defined as working 35 hours or more per week.

[9] For example if one worker retired in June 1990, then in the data we observe that he completely retired from the labor market in 1991 when he is first observed not working.

ratio is nearly the same at the weekly level (Table 1).[10] Figure 2 and Figure 3 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.

Figure 2: Labor supply profiles, sharp v.s. smooth retirement, from CPS [left], PSID [middle] or HRS [right]

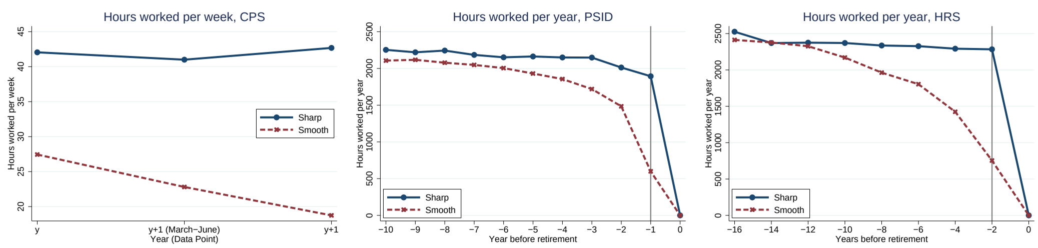


Figure 3: [HRS] Hours worked per week [left] and weeks worked per year [right].

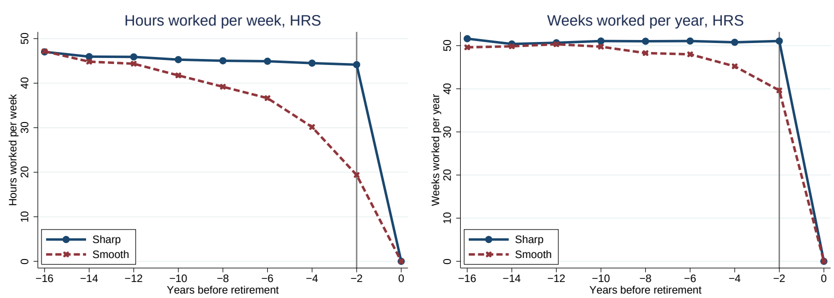


Table 1 summarizes the ratios of individuals working full-time at different interview points before retirement. Many individuals are never observed working part-time before retirement, regardless of marital status, work type, or health status; even among self-employed workers, 30%-75% of them choose sharp retirement.

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

The findings in this article 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. The next section gives a brief review of those approaches.

3 Previous Approaches for Sharp Retirement

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

At old ages, workers retire when the reservation wage exceeds the current wage, which can occur due to an increase in the reservation wage (Gustman and Steinmeier, 1986; Blau, 2008), a decrease in the wage, or both. In most cases, changes in these wages are assumed to be gradual. Even though an abrupt decline in productivity can be caused by a bad health shock, Blau and Shvydko (2011) find that most retirement in the HRS data is not associated with such deterioration in health. Table 1 also shows sharp retirement dominates regardless of health status.

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.

3.1 Social Security or Pension Rules

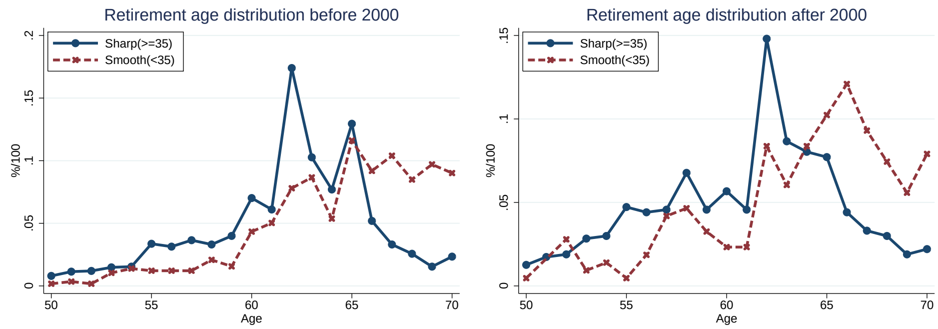
The Social Security policy rules require that individuals are at least 62 years old to receive any benefit, and at least 65 years old to receive full benefit, as well as Medicare. In a model without saving, Rust and Phelan (1997) argue that individuals with low incomes would have to work hard to finance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

The Social Security earnings test might reduce the wage dramatically. The earnings test, however, is just a delayed payment of Social Security benefits, and 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. Since 2000, the earnings test is eliminated after reaching the Normal Retirement Age. 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.

Furthermore, sharp retirement could happen at any age (Figure 4). Similar retirement age distributions are found in PSID and HRS data. It is clear that the Social Security policy is not enough to explain sharp retirement.

On the other hand, private pensions are only able to explain a small portion of sharp retirement. Similar to Social Security, Defined benefit (DB) pension benefits are illiquid before a certain age (usually 55, 60 or 62), otherwise a strong benefit-deduction penalty applies. However, Figure 4 shows that sharp retirement could happen at any age. Defined contribution (DC) pensions mainly depend on the accumulated account and is not directly associated with any specific employer, thus do not necessarily imply sharp

Figure 4: [CPS] Retirement age distribution before the year 2000 [Left] or after [Right].



retirement. 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, and given that Mandatory retirement is prohibited in most jobs since 1994, a standard model implies that one individual would gradually reduce labor supply, implying smooth retirement. To explain sharp retirement, two closely related approaches are used in the literature, namely, discrete labor choices and fixed costs.

3.2 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. Please refer to Hurd (1996) for a rather complete review. 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. I also find that workers who switch occupations are indeed less likely to retire sharply than those who do not. This is in line with the literature on bridge jobs (e.g., Ruhm, 1990) and occupation changes (e.g., Gilleskie and Hoffman, 2014).[11]

However, there are many sharp retirements for which labor market rigidity is of limited explanatory power. First of all, the ratios of sharp retirement are quite high in all occupations and industries. Second, 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. Third, 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 (Table 1).

3.3 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 (Cogan, 1981; Rogerson and Wallenius, 2010). 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; Rogerson and Wallenius, 2013). Such convexity may induce the worker to

[11]Research finds human capital is likely to be occupation specific (e.g., Kambourov and Manovskii, 2009a,b). After switching occupation, the wage is likely to drop, which would induce less labor supply.

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.[12]

However, such fixed costs are far from enough to generate sharp retirement. Cogan (1981) estimates that the average annual fixed monetary cost of work is 28.3% of annual earnings for working, married women. French (2005) estimates that in models without a convex wage-hours mapping[13] fixed time costs are 26%-29% of total leisure endowment, or 53%-59% of a full-time job,[14] which is 4.23-4.7 hours on top of any hours worked in each working day. Even including the convex mapping,[15] which is effectively equivalent to fixed costs, fixed time costs are 1.13-1.37 hours per day. 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 fixed time costs need to be higher than a part-time job in order to create sharp retirement when the intertemporal elasticity of substitution 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

[12]This 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.

[13]Table 2 in French (2005), specification (1) and (2), without accounting for tied wage-hours offers.

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

[15]Table 2 in French (2005), specification (3) and (4), accounting for tied wage-hours offers.

greater detail in Subsection 8.2.

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 the previous subsection, which is possible to explain some, but not all, of sharp retirement behavior.

This article 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 article is to provide an alternative explanation—a work habit model—to explain sharp retirement, presented and estimated in the next three sections.

4 The Work Habit Model

I propose and estimate a life-cycle labor supply model with work habit wherein sharp retirement can be explained by workers quitting "cold turkey." The household head is considered to be the one who makes all relevant decisions regarding consumption, labor supply, and Social Security benefits application. There are two sources of uncertainty each individual faces: survival probability and stochastic wages. Retirement is defined as a labor supply of zero and it is not an absorbing state; workers can re-enter the labor force after retirement.

Time is discrete. At time t, each living individual derives utility from consumption and labor supply,[16]

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

with $\alpha_h > 0$, $\alpha_{hl} > 0$, $\alpha_{hs} > 0$. The first term in (1) is a Constant Relative Risk Aversion (CRRA) utility function in consumption c_t. The second, third, and fourth terms capture disutility of labor supply h_t, where S_t measures the stock of work habit accumulated from working, according to the following process,[17]

$$S_{t+1} = (1-\delta) S_t + h_t, \ \delta \in [0, 1]. \tag{2}$$

The introduction of work habit in the preference is essential in this model, so is discussed in details below.

4.1 Work Habit

There are three terms in the disutility of labor supply in the period utility (1). The first two terms are standard and capture the idea that leisure is a normal good. Note that the second term, $-\alpha_{hl} h_t$, is a variable cost, and therefore is different from the fixed costs used in Cogan (1981), French (2005), Rogerson and Wallenius (2010), or French and Jones (2011), where a fixed amount of costs has to be paid for any positive amount of labor supply. It is also different from Donald and Hamermesh (2009). In their model participation in the labor market reduces efficiency of home production or leisure but the

[16] Subscript i for the individual is suppressed to save notation.

[17] The way that an individual accumulates work habit is essentially similar as the learning-by-doing process estimated in Imai and Keane (2004). The difference is that work 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. Due to computational complexity introduced by the learning-by-doing this excise is not conducted in this paper. The wage process used in this paper is a function of age or experience and is assumed to be exogenous.

degree of efficiency lost is independent of the amount of labor supply as long as it is positive. On the other hand, in this model the linear disutility term is proportional to labor supply, which makes it more similar as a "tax" on the wage. Technically, it generates a positive marginal disutility of working at zero labor supply implying a positive reservation wage. Without this term, the marginal disutility of labor supply at zero is non-positive, which generates a non-positive reservation wage and therefore nobody ever retires.

The last term, $\alpha_{hs} h_t S_t$ with $\alpha_{hs} > 0$, represents that work habit affects the disutility of labor supply in a form of adjacent complementarity, which affects labor supply in both the intensive and the extensive margins.

Given the same parameter γ, work habit reduces the effective elasticity of labor supply on the intensive margin. The first order condition for an interior solution of labor supply h_t is

$$w_t c_t^{-\rho} = \alpha_h h_t^{\frac{1}{\gamma}} + \alpha_{hl} - \alpha_{hs} S_t - \beta V_S\left(a_{t+1}, S_{t+1}\right), \tag{3}$$

where $V\left(a_t, S_t\right)$ is the value function, $V\left(a_t, S_t\right) = \max_{c_t, h_t} U\left(c_t, h_t\right) + \beta V\left(a_{t+1}, S_{t+1}\right)$, and $V_S\left(a_{t+1}, S_{t+1}\right) = \frac{\partial V(a_{t+1}, S_{t+1})}{\partial S_{t+1}}$. The left hand side is the marginal benefit while the right hand side is the marginal disutility of labor supply. If there is no work habit S_t, then we have

$$\ln h_t = \gamma \ln\left(w_t c_t^{-\rho} - \alpha_{hl}\right) - \gamma \ln \alpha_h. \tag{4}$$

Therefore, when the wage declines gradually as one ages, the optimal labor supply would also decline gradually, generating smooth retirement just as in standard models. The elasticity of labor supply on the intensive margin is roughly γ if α_{hl} is relatively small.

However, the presence of work habit S_t, which is accumulated past labor supply, reduces the marginal disutility of current labor supply. That is, the more one worked in the past (the higher S_t), the more one wants to work today (the higher h_t); labor supply is complementary among adjacent periods.

This makes the labor supply less sensitive to the wage change and therefore reduces the elasticity of labor supply on the intensive margin. Note that $V_S(a_{t+1}, S_{t+1}) = \alpha_{hs} h_{t+1} + \beta(1-\delta) V_S(a_{t+2}, S_{t+2})$. For demonstration, further assume $\delta = 1$ and rewrite the first order condition (3) as

$$h_t^{\frac{1}{\gamma}} \left(\alpha_h - \beta \alpha_{hs} h_{t+1} h_t^{-\frac{1}{\gamma}} \right) = w_t c_t^{-\rho} - \alpha_{hl} + \alpha_{hs} S_t \tag{5}$$

or

$$\ln h_t + \gamma \ln \left(1 - \beta \frac{\alpha_{hs}}{\alpha_h} h_{t+1} h_t^{-\frac{1}{\gamma}} \right) = \gamma \ln \left(w_t c_t^{-\rho} - \alpha_{hl} + \alpha_{hs} S_t \right) - \gamma \ln \alpha_h. \tag{6}$$

The second term in the left hand side is an increasing function of h_t.[18] Therefore an infinitesimal increase of wage $\Delta \ln w_t$ will cause a less than $\gamma \Delta \ln w_t$ increase in $\ln h_t$. That is, the elasticity of labor supply on the intensive margin is smaller than γ. This pattern is very similar to the human capital model in Imai and Keane, 2004.

On the other hand, work habit may induce a quick and discontinuous drop of labor supply on the extensive margin, which would lead to retiring cold turkey.[19] Recall that in standard labor supply models, the concave utility function and convex constraints imply graduate decline in labor supply when wages decrease gradually with age. However, with the inclusion of adjacent complementarity in labor supply, the utility function (1) is not jointly concave in h and S, since the Hessian matrix is not negative semidefinite. In such case, the corner solution of labor supply becomes relevant. This is best illustrated in the following example.

Figure 5 shows a simplified model with a linear wage profile from 5 in Periods 30-34 to 0 at the last five periods. Assume no Social Security benefits and the work habit fully depreciates after one period ($\delta = 1$). The model is solved for initial values of $A_{30} = 0$ and $S_{30} = 0.4$. The life-cycle

---

[18] h_{t+1} is a non-decreasing function of h_t, but $\frac{\partial \ln h_{t+1}}{\partial \ln h_t} < 1 < \frac{1}{\gamma}$ around a stable "steady state." Suppose $\frac{\partial \ln h_{t+1}}{\partial \ln h_t} > 1$, then a small increase in h_t will cause h_{t+j} to explode.

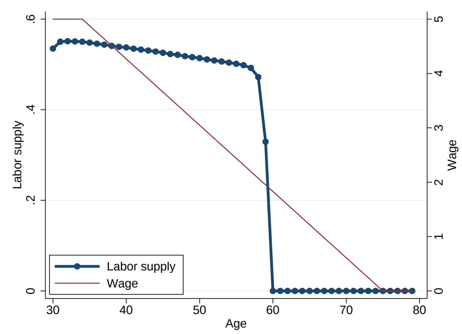
[19] It is argued in Becker and Murphy (1988) that "cold turkey" is the only way to end a strong habit in a similar setup with an infinite horizon.

labor supply profile is simulated. It is shown in Figure 5 that the labor supply is highly inelastic on the intensive margin but extremely elastic on the extensive margin. During Periods $58-60$, the simulated individual tries to reduce labor supply as quickly as possible and retires sharply. In this example, the first order condition of labor supply is

$$\alpha_h h_t^{\frac{1}{\gamma}} + \alpha_{hl} = \alpha_{hs} S_t + \beta \alpha_{hs} h_{t+1} + w_t c_t^{-\rho}. \tag{7}$$

At Period 60, the labor supply calculated from this first order condition is $\hat{h}_{60} = 0.338$ (given $h_{61} = 0$). However, this interior solution is not optimal due to the nonconcavity, and the optimal labor supply is the corner solution $h_{60} = 0$, solved numerically in this example.

Figure 5: A simplified example of the work habit model.



This example also shows that the adjacent complementarity causes a fast and steep decline of labor supply upon retirement. Notice that there are two components contributing to the complementarity in the marginal benefit part of the first order condition (7), $\alpha_{hs} S_t$ and $\beta \alpha_{hs} h_{t+1}$. The first part is the effect of adjacent past labor supply while the second part is the effect of adjacent future labor supply. If one anticipates to retire next period, then at the current period the second effect disappears and therefore results a steep decline of labor supply. In this example, at Period 59, the labor supply calculated from (7) is $h_{59} = 0.329$, given $h_{60} = 0$, while it is $h_{58} = 0.472$ at Period 58, $h_{57} = 0.492$ at Period 57, and $h_{35} = 0.548$ at Period 35. Even though the wage is linearly declining since Period 35, the decline in the labor supply is relatively small from Period 35 to Period 58 (dropped 0.076 or 13.9% over 23 periods). In contrast, the labor supply drops 0.143 or 30.3% over one period from Period 58 to 59, and drops 100% from Period 59 to 60. The steep decline of labor supply at Period 59 is caused by that this individual retires at Period 60 and therefore the second half of the complementarity effect disappears ($h_{t+1} = 0$ for $t = 59$ in the first order condition). The complete decline of labor supply at Period 60 is from that the corner solution dominates the interior solution due to nonconcavity. In other words, it is like a "chain reaction" when rational individuals react to their expected retirement.

It is less satisfying that the model has a "bridge" point of Period 59 rather than a completely sharp retirement. However, the point is that this steep decline of labor supply at the bridge point happens very quickly—one period in this example. If a period is defined as a smaller grid, for instance a month, then the "bridge" point would be just one or several months. The resulting retirement would still appear quite sharp at the yearly level. In contrast, those workers who choose smooth retirement gradually reduce labor supply over a ten-year period (Figure 2).

Work habit can be interpreted as working routines, which could be physical or mental. 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, or teaching at the beginning of a semester seems harder than later.

Work habit can also be interpreted as the status in a social group, at least for some occupations. It could be relationship with colleagues, or other psychological status in the relationship or hierarchy. Nevertheless, building up and enjoying such status require one's presence in previous periods.

This adjacent complementarity is 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 article applies and estimates intertemporal nonseparability on the extensive margin.

4.2 Other Constraints

The remainder of the model is as follows. Details about Social Security rules and tax codes are laid out in Appendix B and C.

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, and it is assumed that a qualified application is always granted.

Each individual faces a budget constraint

$$A_{t+1} = (1+r)A_t + Y_t\left(\max\left\{0, w_t\left(h_t - \tau\right)\right\}, y_t^f, ss_t\right) - c_t + tr_t, \tag{8}$$

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 $w_t\left(h_t - \tau\right)$, other income y_t^f, Social Security benefits ss_t if applicable, and the tax code. The fixed time costs, τ, has to be paid if the individual works $(h_t > 0)$. If eligible, individuals receive government transfer, tr_t. Individuals are not allowed to borrow against future income. This implies a non-negative borrowing constraint, $A_{t+1} \geq 0$.

For a household with more than one member, the adult-equivalent consumption $\hat{c}_t$ is calculated according to the formula $\frac{\hat{c}_t}{c_t} = (\#adults + a_1 \times \#kids)^{a_2}$, where $\#adults$ is the number of adults and $\#kids$ is the number of children in the household. Values of $a_1 = 0.70$ and $a_2 = 0.75$ are used in the paper following Citro and Michael (1995), page 178.

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 estimated directly from the data.

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

$$\ln w_t = x_t^w \pi_w + \xi_t, \tag{9}$$

$$\xi_t = \phi_w \xi_{t-1} + \epsilon_t^w,\ \epsilon_t^w \sim \mathcal{N}\left(0, \sigma_w^2\right). \tag{10}$$

where x_t^w includes age and its square. The autoregressive component has a correlation coefficient ϕ_w and an independent, identically distributed innovation ϵ_t^w drawn from a normal distribution.

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 less expenses. In particular, for this article, 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 and its square,

$$y_t^f = x_t^f \pi_f \tag{11}$$

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

$$tr_t = max\left\{0, c_{min} - ((1+r)A_t + Y_t)\right\}. \tag{12}$$

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

There is a bequest motivation in the form of

$$b(A_t) = b_1 \frac{(b_2 + A_t)^{1-\rho}}{1-\rho} \tag{13}$$

where b_1 captures the relative weight of the bequest motivation and b_2 determines its curvature (DeNardi, 2004).

There are other important factors not included in the model. For instance, health and health insurance are important factors that individuals take into account when they decide the retirement timing (e.g., Rust and Phelan, 1997; Blau and Gilleskie, 2001, 2008; French and Jones, 2011). However, since I am focusing on how individuals retire rather than why they retire, I decide not to include them in the model. This compromise to increasing tractability will certainly include some bias in the estimation, mostly in the timing of retirement. Similarly, ignoring the complication of private pension rules reduces the computational burden at the cost of possible overestimation of the degree of adjacent complementarity; ignoring stochastic medical expenses comes at a cost of overestimating the bequest motivation.

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

$$\begin{aligned} V(X_t) &= \max_{B_t, h_t, c_t} \frac{c_t^{1-\rho}}{1-\rho} - \alpha_h \frac{h_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} + \alpha_{hs} h_t S_t - \alpha_s S_t \\ &\quad + \beta \left\{ (1 - q_s(t))\, b(A_{t+1}) + q_s(t) \int V(X_{t+1})\, dF(X_{t+1} | X_t, c_t, h_t, B_t) \right\} \end{aligned} \tag{14}$$

subject to the law of motion for work habit (2), the budget constraint (8), the non-negative borrowing constraint, the wage process (9)-(10), the other income process (11), the government transfers (12), as well as the bequest motivation (13). X_t is the vector of state variables, $X_t = (A_t, S_t, AIME_t, \xi_t, B_{t-1})$.

4.3 Model Solution

The model is solved numerically. At time t, the state space is discretized into a finite grid space: log-linearly into 41 points, respectively, in asset A_t and $AIME_t$, and linearly into 47 points in work habit S_t. The autoregressive component in the wage process is discretized into a discrete Markov process with 9 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 $1,422,126$ points in each period.

The value function and decision rules are solved backwards. At each period, the value function is solved as shown in (14). 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 and consumption c_t. Linear interpolation is used in calculating the next-period value function and policy functions at any state value.

5 Data, Calibration and Estimation

5.1 Data and Calibration

I use the Health and Retirement Study (HRS) data to estimate the parameters of interest. Compared with the PSID data, the HRS data collects more detailed and accurate information relevant to this article, such as Social Security income, asset, and labor supply. 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 article uses nine waves of data, through 2008, when the cohort was aged 67-77. The descriptive statistics for the initial conditions are presented in Table 2 (asset, AIME and wage in 2004 dollars). The groups of sharp retirement, smooth retirement, or not retired are according to their retirement behavior within the sample period from 1992 to 2008.

Table 2: Descriptive statistics of initial conditions.

	Overall		Sharp retirement		Smooth retirement		Not retired	
	Mean	*S.E.*	*Mean*	*S.E.*	*Mean*	*S.E.*	*Mean*	*S.E.*
N	1,506		708		254		544	
Age	57.6	(3.45)	57.8	(3.49)	58.7	(3.44)	56.8	(3.21)
Asset	165,001	(185,683)	160,492	(167,730)	186,907	(186,344)	160,640	(206,112)
AIME	2,913	(1,837)	2,985	(1,893)	2,727	(1,568)	2,906	(1,877)
Wage	21.50	(12.19)	21.98	(11.55)	20.09	(12.20)	21.42	(12.95)
Labor supply	2,230	(766)	2,269	(645)	1,792	(937)	2,384	(748)

I assume that each individual can live as long as 84 years. This seems arbitrary but reasons are twofold. First, this article concerns retirement behavior, not retiree behavior. Of course the 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 article 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 workers who choose smooth retirement, on average the smoothing process spans ten years (Figure 2). 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 age 50 which is defined as period 50 as well, and ends at age 84 which is defined as period 67, and runs every two years which is equivalent to one period. Therefore, the early retirement age is at period 56 (or age 62) and the normal retirement age is at period 58 (or age 66).

The risk free interest rate r is set at 0.06. In the model the total time endowment at each period is defined as one unit. The total time endowment in the HRS data is defined as $5,600$ hours per year (16 hours per day for 350 days, excluding 15 public holidays). When calculating data moments, I divide yearly labor supply and asset in the data by $5,600$ to translate them into those in the model. The consumption floor is set at 0.78 after French and Jones (2011).

5.2 Estimation

The model is estimated in two steps. In the first step, I estimate data generating processes directly from the data, which includes survival probabilities $q_s(t)$ and the other income process π_f. The second step is to estimate all remaining parameters (as listed in Table 3), including the wage process using the method of simulated moments (MSM). The estimation procedure is described in Appendix D.1.

Two sets of moment conditions at each age are chosen to represent the life-cycle profiles. One set is calculated from all individuals while the other set is calculated from individuals choosing sharp retirement. Each set includes the first moments of assets, logarithm of wages, labor force participation rate, hours worked, and labor force re-entering rate, as well as the second moment of wages. The moment conditions also include the ratios of individuals who choose sharp retirement or smooth retirement at each age. In total there are 136 moment conditions. The details of moments conditions are described in Appendix D.2.

In general, most parameters are identified jointly by moments listed above. The identification of specific parameters will be discussed along with the estimation results.

5.3 Initial Conditions and Preference Heterogeneity

I consider $10,000$ individuals in the simulation. For each simulated individual, the initial conditions are drawn jointly from the data, including age, asset, AIME and wage.

The unobserved initial value of work habit S_{t_0} at the initial period t_0 is assumed to be a linear function of the AIME and the labor supply at the first observed period, $S_{t_0} = \kappa_0 + \kappa_1 \cdot h_{t_0} + \kappa_2 \cdot AIME_{t_0}$. This is to assume that the AIME which captures the past work experience and the initially observed hours worked which captures the current work status are good indicators of the accumulated level of work habit.[20] I also tried assuming S_{t_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. 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 $\left\{\alpha_{hl}^j, \alpha_{hs}^j\right\}$, $j = 1, 2$. For each individual the probability of having $\left\{\alpha_{hl}^1, \alpha_{hs}^1\right\}$ is modeled as a logistic function of a subset of initial state variables,[21]

$$\Pr\left(j = 1\right) = \frac{1}{1 + \exp\left(-\left(\kappa_3 + \kappa_4 \ln w_{t_0} + \kappa_5 AIME_{t_0} + \kappa_6 A_{t_0}\right)\right)}. \tag{15}$$

Assuming such heterogeneity is mainly to improve the model fit, and is

[20]The AIME and the labor supply at the *first* observed period are exogenous to the worker as they are already given when the model begins. The AIME and the labor supply in all subsequent periods are endogenous.

[21]Using more than two types yields very similar results.

different from pre-labeling workers according to their observed retirement behavior. As discussed in Section 6, not all individuals with same type of preference choose the same retirement behavior.

6 Baseline Results

The parameter estimates are presented in Table 3 with standard errors reported in parentheses.

Table 3: Parameter estimates in the baseline model (standard errors in parentheses)

Parameters		Estimates	Parameters		Estimates
Habit	$\alpha_{hs,1}$	2.112 (0.0032)	Labor supply	γ	0.434 (0.0016)
	$\alpha_{hs,2}$	0.019 (0.028)		α_h	20.749 (0.031)
	δ	0.912 (0.0021)		$\alpha_{hl,1}$	1.010 (0.0025)
Prob($j = 1$)				$\alpha_{hl,2}$	0.540 (0.0737)
(constant)	κ_3	0.422 (0.063)	Time discount	β	0.996 (0.0006)
($\ln w_0$)	κ_4	0.217 (0.052)	CRRA	ρ	2.267 (0.0002)
($AIME_0$)	κ_5	$5.600e-3$ (0.0061)	Wage: AR(1)	ϕ_w	0.943 (0.0006)
(A_0)	κ_6	$5.865e-3$ (0.0054)	Wage: error	σ_w	0.360 (0.0008)
Initial S			Wage: age	π_w	$8.649e-3$ (0.0029)
(constant)	κ_0	0.224 (0.0021)	Wage: age^2		$-1.116e-3$ (0.0015)
(h_0)	κ_1	0.011 (0.0033)	Wage: $cons$		3.775 (0.0016)
($AIME_0$)	κ_2	0.014 (0.0042)	Bequest weight	b_1	572.263 (5.216)
fixed time costs	τ	0.028 (0.0011)	Bequest shifter	b_2	8.918 (0.096)

One of the parameters of primary interest is the level of adjacent complementarity in labor supply, α_{hs}. Its assumed heterogeneity is confirmed in the data; two different values of α_{hs} are estimated. These parameters, together with the parameters directly related to labor supply, are identified by the moments of labor supply and the proportions of sharp and smooth retirement. The first type has much stronger adjacent complementarity ($\alpha_{hs,1} = 2.112$) than the second type ($\alpha_{hs,2} = 0.019$). Among individuals with strong adjacent complementarity, about three quarters take sharp retirement. Among the remaining one quarter, nearly half of them work half time before retiring completely. This is mainly caused by a very smooth decline in the hourly wage process. The other one-eighth individuals either retire at the beginning or never retire during the sample period. Notice that the linear disutility of labor supply is also higher for this group of individuals, which helps generate a positive marginal disutility of working at zero labor supply.

Individuals with weak adjacent complementarity are basically similar to those in the neoclassical life-cycle labor supply model, with $\alpha_{hs,2}$ not significantly different from zero. Among this group of individuals, only one-fifth retire cold turkey, mostly in response to a steep decline in hourly wages. This verifies that a neoclassical model is not enough to generate sharp retirement observed in the data.

The estimated fixed time costs are 0.028, equivalent to 157 hours per year. This is comparable to the average commute time in PSID and in the American Time Use Survey 2003 (both reports around 160 hours per year), but is much smaller than other estimates in previous literature. Such discrepancy comes from different mechanisms generating sharp retirement. French (2005) and French and Jones (2011) estimate the fixed time costs of working mostly higher than $1,000$ hours per year. Rogerson and Wallenius (2010) calculate that fixed time costs need to be higher than a part-time job in order to induce sharp retirement. In this article, however, the major driving force of sharp retirement is the cold turkey quitting behavior (to break the work habit). Therefore, the estimated fixed time costs are much smaller in this model.

The identification of such fixed time costs comes from the labor force re-entry rates. Notice that the labor force re-entry rate is the transition rate from non-participation to participation. Given fluctuation in the wage process, high fixed time costs and weak adjacent complementarity lead to more frequent retire-and-re-enter to the labor force, compared to low fixed time costs and strong adjacent complementarity. This is because in the latter case in order to take advantage of the strong adjacent complementarity workers would like to avoid losing and re-accumulating work habit. Therefore the moments of labor force re-entry rates on the whole sample as well as on the group of sharp retirement help pin down the estimate of fixed time costs.

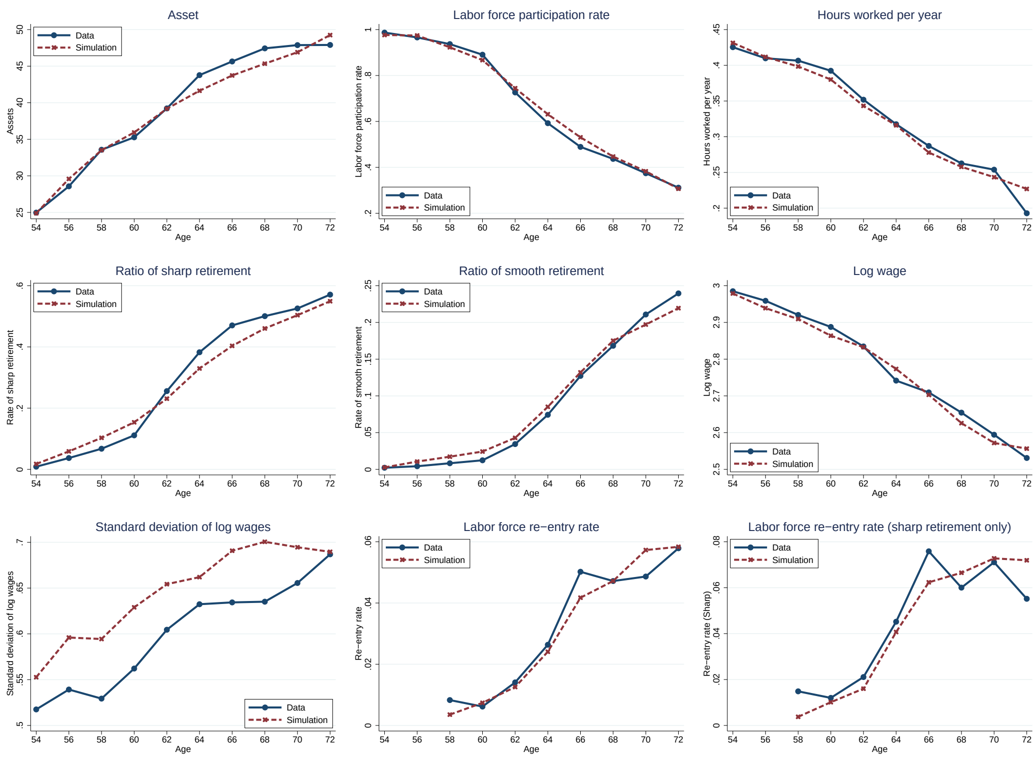
The estimated parameter in the convex disutility of labor supply is $\gamma = 0.434$. However, this is different from the intertemporal elasticity of substitution parameter in a standard model, due to the intertemporal nonseparabilities. As discussed in Subsection 8.1, the simulated micro elasticities of labor supply range from .022 to .056 while macro elasticities are higher, ranging from 0.132 at age 56 to 0.771 at age 68.

The time discount factor, β, is identified by the survival probabilities. For this reason, as in French (2005), my estimator is larger than most estimates in previous literature. The coefficient of relative risk aversion in consumption is 2.267, similar to estimates in French (2005). The bequest weight—572.263—is much larger than the 0.0223 estimated in French and Jones (2011). As discussed in Section 5, this is expected because the maximum age is only 84 versus 94 in that paper and medical expenses are not included in this model. The high saving rates with a high bequest weight predicted by my model are consistent with the findings in Scholz et al. (2006).

Figure 6 shows that the model is able to replicate assets, the labor force participation rates, and the hours worked within the relevant period. The model generates slightly less sharp retirement and more smooth retirement in the simulation than in the data. This is potential evidence that other factors besides work habit and fixed costs are also important in inducing sharp retirement, such as some level of labor market rigidity. The model is able to fit the accepted wages but over-predicts their variation. This is due to the lack of stochastic shocks in the model. It only has the mortality shock and the wage shock, but no health, medical expenses, or other income shocks. Therefore the second moment of the wages is overpredicted, trying to match other moments with the data. The model generates fairly good predictions for the labor force re-entry rates for both the whole sample and the subsample of individuals who take sharp retirement.

To further check the fit of the model, I plot the labor supply profiles relative to the first retirement wave in Figure 7, which are not directly forced to match the data. The model fits well in hours for individuals choosing smooth retirement, and fairly well for individuals choosing sharp retire-

Figure 6: Model Fit: data and simulation

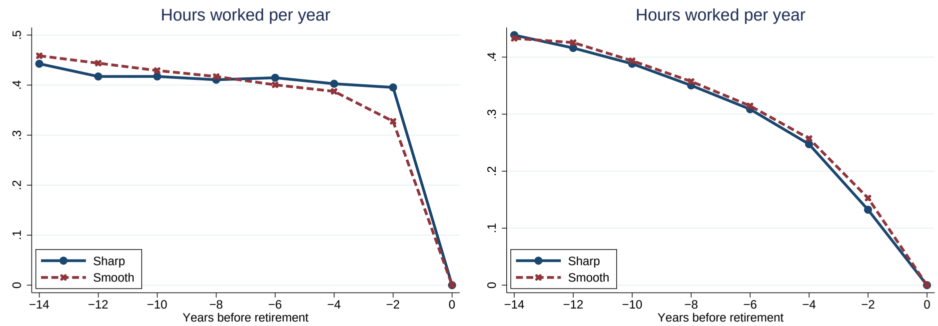


ment. Notice that two years before the first retirement wave, the model predicts a smaller hours worked than the data. This is due to the “bridge” point discussed in the example in Subsection 4.1. Comparing with that particular example which is a fully deterministic case, the uncertainty in the wage process in the full model here helps make the bridge point less obvious.

7 Experiments

Policymakers are interested in how changes of social security policies will affect individual behavior. In this section I conduct three counterfactual policy experiments. The first experiment sets the Early Retirement Age (ERA) one period later, which is equivalent to changing ERA from age 62 to 64. The

Figure 7: Labor supply of workers choosing sharp [left] or smooth [right] retirement.



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 4.

Table 4: 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

Note: 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.)

The first column of Table 4 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. Thus shifting the ERA affects individuals choosing sharp retirement more significantly.

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, delay retirement by 0.57 months and apply Social Security benefits 2.40 months earlier, all per lifetime. 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.

8 Discussion

In this section I show that the work habit model generates different elasticities of labor supply on the intensive and extensive margins. I also compare this model with models of fixed costs, arguing that it fits the data better.

8.1 Labor Supply Elasticities

Besides rationalizing sharp retirement, the work 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. 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 and to respond to wage changes less sensitively. In the example discussed in Subsection 4.1, up to Period 57 the first order condition 7 is roughly $\alpha_h h_t^{\frac{1}{\gamma}} = 4.723 + 0.709 w_t - \alpha_{hl}$, The first term in the right hand side comes from the adjacent complementarity and dominates the wage incentive which is the second term. Therefore it generates a rather small elasticity of labor supply on the intensive margin. The estimated labor supply elasticity on the intensive margin from this example is 0.127, smaller than the elasticity parameter $\gamma = 0.2$.

From the baseline model, I estimate labor supply elasticities at different ages and results 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, while the elasticity on the extensive margin, $\epsilon_{e,t_2}^{t_1}$, is calculated by the change in the labor force participation rates.

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

Elasticity	Macro		Micro (intensive margin)		Extensive margin	
Age t	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	0.135	0.020	0.033	0.057	0.101
60	0.072	0.251	0.025	0.045	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	0.690	0.026	0.031	0.242	0.658
68	0.163	0.771	0.001	0.028	0.162	0.742

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 article. Table 5 shows that the micro elasticity is quite small, mostly ranging from .022 to .056. The model also generates relatively larger elasticities on the extensive margin, from 0.101 to 0.742. The macro elasticities range from 0.132 at age 56 to 0.771 at age 68.

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 almost 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.

8.2 Comparing with Fixed Costs

This section derives and compares the implications of the work habit model with 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.[22] For instance, within a period of one year, working 40 hours per week for 20 weeks is more attractive than working 20 hours per week for 40 weeks since the former pays half of the fixed costs of the latter.[23] 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.

[22]Some literature assumes a re-entry cost but usually is 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.

[23]There are fixed costs that accrue on a longer time horizon than week, for instance an extra car for commute or health/life insurance policy. However they don't affect our example here since they are like "sunk" costs.

Figure 3 shows that individuals who take smooth retirement in the HRS data are more likely to reduce hours worked per week rather than weeks worked per year. 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 work habit model the effect of adjacent complementarity induces the individual to smooth labor supply within the period whenever it is optimal to work. This implication of smoothing labor supply is on the intensive margin and is different from the cold turkey quitting which is on the extensive margin. In the work habit model, when one retires, he most likely quits cold turkey if the level of adjacent complementarity is strong, otherwise he quits smoothly. However, regardless of the heterogeneity in the level of adjacent complementarity, when one works, he wants to smooth the work habit through a continuous spell of working periods. This is illustrated in the example in Figure 5.

Some might argue it is possible that individuals might cluster labor supply at the daily level within each week by cutting back days worked per week instead of hours worked per day. Unfortunately, the three data sets used in this article 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 commute 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 variable and the model is similar to standard models, which find difficulty in justifying sharp retirement.

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 article, 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 labor production function. This article proposes and estimates a model incorporating a different approach, which essentially assumes a non-concave preference with habit persistence. I argue that this is a reasonable assumption and well-supported by previous literature and various datasets. Using HRS data, I estimate a life-cycle labor supply model with work habit in which sharp retirement can be explained by workers retiring cold turkey to quit work habit. The model produces reasonable parameter estimates with good model fits, including a more empirically plausible estimate of fixed time costs.

This model provides an alternative explanation to rationalize sharp retirement, which is supported better by data than existing models. In particular, I show that, for individuals taking smooth retirement, on average they 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.

A Data Appendix

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 (DP1-DP4) for each individual (Table 6). MORG CPS collects labor market participation information (labor force status, hours, hourly or weekly earnings) for last week while March CPS collects those information for last year. March CPS collects yearly earnings instead of hourly or 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. There are totally $3,309$ individuals, as shown in Table 7.

Table 6: 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	*March*	*4*	*June*	*5*	*March*	*8*	*June*
2	*March*	4	*May*	6	*March*	8	*May*
3	*March*	4	*April*	7	*March*	8	*April*
4	*March*	4	*March*	8	*March*	8	*March*

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

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 de-

Table 7: 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

mographic information with substantial detail on income, employment and family structure. From 1968 to 1997 individuals in PSID are interviewed and re-interviewed every year. Since 1997, the survey is conducted biennially.

In this article, 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.[24]

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, eliciting information about demographics, income, assets, job status and history, family structure and many others. This article 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, when respondents were aged 67-77. This article primarily uses RAND HRS data, version K.[25]

[24]Not all individuals enter the sample from 1968.

[25]"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."

The following sample selection rules are applied in order:

(1) Exclude individuals with top 2% asset/wage or more, or with bottom 2% asset/wage or less at any point.

(2) White non-Hispanic male only.

(3) Exclude observations ever applied or receiving Supplemental Security Income (SSI) or Social Security disability SSDI, or ever disabled in any wave.[26]

(4) 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).

The final sample consists $1,506$ individuals with $9,802$ observations.

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, and the Medicare tax which is 1.45%, uncapped. I use the Social Security tax limit in 2004 which is $87,900.

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.

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 8.

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

Table 8: 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	$\geq 329,351$	$229,371.63 + 0.6355\,(Y - 329,350)$

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), which varies by the year of birth as listed in Table 9. At the NRA, the worker receives retirement benefits equal to the Primary Insurance Amount (PIA), which is a function of Average Indexed Monthly Earnings (AIME),

$$\begin{aligned} PIA &= 0.9 * min\left\{bp_1, AIME\right\} + 0.32 * min\left\{bp_2 - bp_1, max\left\{0, AIME - bp_1\right\}\right\} \\ &\quad +0.15 * max\left\{0, AIME - bp_2\right\} \end{aligned} \tag{16}$$

The bend points bp_1, bp_2 vary by year of entitlement as listed in Table 10. For simplicity, I use the 2004 bend points $(bp_1,\ bp_2) = (612,\ 3689)$ for all the individuals in my sample.

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 wage inflation by multiplying the wage growth rate relative to the base year, which is at age

60 (two years prior to the year of first eligibility). The wage growth 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 are not adjusted. Interestingly, the wage growth rate of the national average wage index is very similar to the growth rate of CPI-U after Year 1969 (Figure 8), 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, which is computationally infeasible. Instead I apply an approximating method, taking into account the wage growth pattern over the lifecycle

$$\begin{aligned} AIME_{t+1} &= f_t^a\left(AIME_t, I_t, L_t\right) \quad (17) \\ &= AIME_t + max\left\{0,\ \frac{sse_t}{35\times 12} - 1\left\{tw \geq 35\right\}\cdot share_{min}\cdot AIME_t\right\} \end{aligned}$$

where $sse_t = min\left\{wh_t,\ s\bar{s}e\right\}$ 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 9 lists the estimated $share_{min}$ from CPS data for age 52 to 76, assuming the starting working age of 16.

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,

$$\begin{aligned} AIME &= min\left\{bp_1, \frac{PIA}{0.9}\right\} + min\left\{bp_2 - bp_1, max\left\{0, \frac{PIA - 0.9 * bp_1}{0.32}\right\}\right\} \\ &\quad + max\left\{0, \frac{PIA - (0.9 * bp_1 + 0.32 * (bp_2 - bp_1))}{0.15}\right\} \\ &= \frac{min\left\{550.8, PIA\right\}}{0.9} + \frac{min\left\{984.64, max\left\{0, PIA - 550.8\right\}\right\}}{0.32} \\ &\quad + \frac{max\left\{0, PIA - 1535.44\right\}}{0.15} \quad (18) \end{aligned}$$

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 reduced 5/12 of one percent per month. This is summarized in Table 11.

C.3 Delayed Retirement

Delayed retirement after the NRA increases benefits. The delayed retirement credit (DRC) listed in Table 12 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 13. 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 $1 benefit is withheld for every $3 earnings in excess of the higher amount for 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 12. 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.[27]

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. I ignore the earnings test for workers aged 65 or older before the year of 2000, which likely leads to overestimating the effect of work 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.

C.5 Taxable Social Security Benefits

I use the year 2004 rule of taxation on Social Security benefits to all workers.

The Social Security benefits are not taxable if it is the only income. If there is other income, compute "total income" as the sum of half of the benefits and all other income. If total income is no more than the base amount ($25,000 for head of household) then no benefits are taxable. If total income is higher than $34,000 then up to 85% of the benefits could be taxable.

Assume the Social Security benefits are y_{ss} and all the other income is y_o, the taxable part of Social Security benefits is calculated as

$$y_{ss,taxable} = \begin{cases} 0, & \text{if } y_o = 0, \\ 0, & \text{if } y_o + \frac{y_{ss}}{2} \leq 25000 \\ \min\{0.85y_{ss}, \frac{1}{2}\min\left\{y_{ss}, y_o + \frac{y_{ss}}{2} - 25000, 9000\right\}, & \\ \qquad +0.85\max\left\{0, y_o + \frac{y_{ss}}{2} - 34000\right\}\} & \text{otherwise.} \end{cases} \tag{19}$$

[27] In 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).

Figure 8: Relative (to Year 2004) indices of National Average Wage Index and CPI-U.

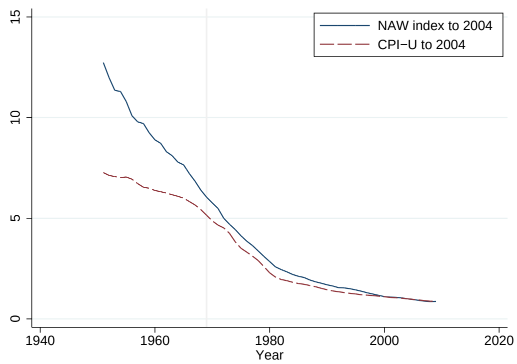


Figure 9: Share of minimum wage on AIME, assuming starting working from age 16. CPS data.

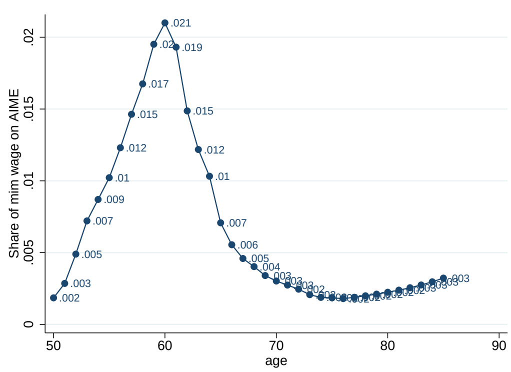


Table 9: Normal Retirement Age (NRA).

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		

Table 10: 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			

D Method of Simulated Moments

D.1 Estimation Procedure

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

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 sur-

Table 11: 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 12: 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%	$\geq$1943	8.0%
1933-34	5.5%		

vival 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 (14) as well as the policy functions.

(6) Generate the life-cycle profile for each simulated individual given the initial condition.

(7) Calculate the simulated moments, and compute the distance between the simulated moments and the data moments.

(8) Pick another set of parameters, repeat steps (5)-(7) until a minimum distance is found.

Table 13: 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.

D.2 Moment Conditions

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

The first set of moment conditions is calculated from all individuals:

(i) Assets: the first moment (mean).

(ii) Logarithm of wages: the first and the second moments (mean and standard deviation).

(iii) The labor force participation rate.

(iv) Hours worked: the first moment.

(v) The ratio of individuals who choose sharp retirement or smooth retirement.

(vi) The labor force re-entering rate.

The moment conditions of log wages and hours worked are calculated as follows

$$E\left(\left(Z_{it} - \bar{Z}_t\right) \cdot D_{it}\right) = 0 \tag{20}$$

where $D_{it} = 1$ if an individual is working and $D_{it} = 0$ otherwise in the data, and Z_{it} denotes log wages or hours worked in the data. $\bar{Z}_t$ is the model's prediction of Z_{it}.

In a similar manner, the second set of moment conditions is calculated from individuals choosing sharp retirement:

(vii) Assets: the first moment.

(viii) Logarithm of wages: the first and the second moments.

(IX) The labor force participation rate.

(X) Hours worked: the first moment.

(XI) The labor force re-entering rate.

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 (equivalent to age 72). The job re-entering rates at period 52-53 are zero so these two moments are dropped. This leaves 136 moment conditions in total.

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.

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.

Blau, David M and Donna B Gilleskie, "Retiree health insurance and the labor force behavior of older men in the 1990s," *Review of Economics and Statistics*, 2001, *83* (1), 64–80.

_ **and** _ , "The Role of Retiree Health Insurance in the Employment Behavior of Older Men*," *International Economic Review*, 2008, *49* (2), 475–514.

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.

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.

Gilleskie, Donna and Denise Hoffman, "Health Capital and Human Capital as Explanations for Health-Related Wage Disparities," *Journal of Human Capital*, 2014, *8* (3), pp. 235–279.

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.

Heckman, James and Burton 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 Accumlation," *International Economic Review*, May 2004, *45* (2), 601–641.

Kambourov, Gueorgui and Iourii Manovskii, "Occupational mobility and wage inequality," *The Review of Economic Studies*, 2009, *76* (2), 731–759.

— **and** — , "Occupational specificity of human capital," *International Economic Review*, 2009, *50* (1), 63–115.

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 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.

Meyer, Bruce D., "AssociationLabor 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.

_ **and** _ , "Nonconvexities, Retirement and the Elasticity of Labor Supply," *American Economic Review*, June 2013, *103* (4), 1445–1462.

Ruhm, Christopher J., "Bridge Jobs and Partial Retirement," *Journal of Labor Economics*, 1990, *8* (4), pp. 482–501.

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.

Ryder, Harl E. Jr. and Geoffrey M. Heal, "Optimal Growth with Intertemporally Dependent Preferences," *The Review of Economic Studies*, January 1973, *40* (1), 1–31.

Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun, "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.