PARIS SCHOOL OF ECONOMICS

Delay the claiming of Social Security Benefits, or Die:

a Survival Analysis of the Claiming Behavior of Older Americans

Yarine Fawaz
e-mail : fawaz@pse.ens.fr

This paper examines Social Security benefit claiming behavior, a take-up decision that has been dealt with in the existing literature as mainly influenced by SS wealth and its determinants. I hereby aim at understanding why so many older people in the United States claim at age 62, which is the Early Entitlement Age, while others prefer waiting until the Normal Retirement Age of 65. I focus essentially on the role of liquidity constraint in this rushing/delaying behavior, but also check the impact of a set of socio-demographic variables such as gender, education, or the difference of age between spouses for married individuals.

The purpose of this paper is to explore SS take-up behavior of older Americans, whether they already stopped working or not, since neither decision is a necessary condition for implying the other. To do so, I use the *Health and Retirement Study*, a longitudinal study upon individuals aged 51 to 61 at baseline (that is, in 1992) and their spouses, which has enriched of 8 biennal waves since then. I first estimate probit models on the probability of claiming SS benefits between the wave before the 62nd birthday and the wave following the 62nd birthday.

I go further than these cross-sectional predictions by estimating a survival analysis using the panel dimension of the HRS. Although this kind of econometric analysis seems the most proper to study single-episode spells like the duration in months between eligibility and SS claiming, it remains a widely unexplored technique in previous literature about claiming and retirement, while very complex duration models have been constructed in labor economics for instance to estimate job duration and unemployment spells.

Non-parametric descriptive methods (Kaplan-Meier and Nelson-Aalen estimators) first provide a global picture of the length of spells before claiming Social Security for every eligible person in the sample, and allow deriving the average and median durations sorted by some of the covariates. Then the most widely used semi-parametric model of duration, the Cox model, is implemented to investigate the impact of every covariate of the model on hazard rates on the total sample first, before running separate regressions, stratifying the analysis, and introducing shared frailty in Cox regressions. The results concerning the covariates confirm those found in the cross-sectional analysis, while the several methods used to allow the hazard to differ other than proportionately between early retirees (those who had stopped working before age 62) the others, highlight how different these two groups are in terms of claiming patterns. Those no longer working at age 62 seem to face much stronger incentives to claim Social Security benefits early than the others. This empirical result is worth more investigation and further modeling as existing retirement literature always considers the incentives to claim SS benefits at age *a* for people who still have a job at this age.

The first baby boomers turned 62 in 2008, the magical age at which one can begin to claim Social Security retirement benefits that will be received from that moment until death. Nearly 80 million Americans of the baby boom generation will become eligible for SS benefits over the next two decades, a phenomenon known as America's "silver tsunami". As the ratio of workers to retirees (*i.e.* all persons aged 50 and over who are not in labor force) is projected to fall from over three to one to around two to one in 2030, and growing expenditures on social security will have to be financed by taxes on a smaller number of workers relative to pensioners or by cuts in the amount of social security benefits, the retirement decision, and more precisely the age of retirement, has become a key variable. Hence there is a large literature investigating the determinants of retirement, but the latter focuses on retirement as cessation of work rather than retirement as SS take-up, or deals with SS when looking for its influence on retirement behavior. Yet we believe that the take-up decision deserves as much focus as that of retirement, as there *is* a take-up decision, as in the case of other social insurance programs. In other words this decision is worth being investigated as people do not claim as soon as they are eligible, as a fatality, but may weigh the pros and cons of claiming early (*i.e.* before the normal age of SS take-up). Indeed by doing so they forgo large benefits accruals that could be paid to them and their spouses until their deaths, but on the other hand enjoy these benefits for more years and without waiting any longer.

For example, a "boomer" with a final salary of $75,000 might receive a $1,320 monthly Social Security check if he collects at age 62 (Mahaney-Carlson (2006)); if the same boomer were to delay until age 70, he would get $2,884 monthly, more than twice as much. Yet, most retirees apply for SS benefits early: according to the Social Security Administration, 72% of current recipients receive reduced benefits because they started their benefits prior to their Full Retirement Age.

Early claiming has therefore become a social norm while it has been designed as an exception. As it is part-though an independent piece- of the larger puzzle of early retirement that reduces the period over which households accumulate wealth and increases the period over which they decumulate, it is crucial for policy implications to understand the pattern of SS benefits take-up of older individuals.

The purpose of this paper is to explore SS take-up behavior of older Americans, whether they already stopped working or not, since neither decision is a necessary condition for implying the other. We do so in 3 steps. First, section I reminds some relevant institutional features of the American Social Security system that will be necessary to understand SS claiming pattern, and briefly reviews the existing SS literature. Section 2 provides a cross-sectional analysis of the determinants of claiming early, using the Health and Retirement Study, which is a nationally representative panel survey of persons aged 51 to 61 at baseline and their spouses, especially designed to investigate retirement behavior and its implications on the health, social, and economic status of the aging population in the US. Individuals are observed the wave before and after they become eligible to SS retirement benefits, so that SS early take-up may be observed. Claiming and retirement decisions are investigated jointly for non-retired individuals as the claiming decision of respondents who are not retired at age 62 is likely to be related to their retirement decision. We find very strong differences in the claiming pattern depending on whether the sample is made of already-retired-at-62 individuals, or still-working-at-62 respondents. In section 3 we use the panel dimension of the HRS to estimate survival models, which are particularly well suited to study single-episode spells like the duration in months between eligibility and SS claiming. Non-parametric estimates (Kaplan-Meier and Nelson-Aalen estimators) provide the global shape of survivor and cumulative hazard functions with no covariates. Then we will estimate the most widely used semi-parametric model of duration, the Cox model, to investigate the impact of every covariate of the model on hazard rates on the total sample first, before running separate regressions, stratifying the analysis, and introducing shared frailty in Cox regressions. The results concerning the covariates confirm those found in the cross-sectional analysis, while the several methods used to allow the hazard to differ other than proportionately between early retirees and those who did not retire before age 62, highlight how different these two groups are in terms of claiming patterns.

I. Background

1. Institutional features

A brief overview of the US Social Security system is required to understand the motivation to this analysis. First, SS retirement benefits earned by reason of one's own contributions can be claimed at any age from 62 to 70, subject to an earning test that has become less stringent over time. The calculation of benefits involves four steps. First, a worker's previous earnings are restated in terms of today's wages by indexing past earnings to wage growth. Second, earnings for the highest 35 years are averaged and divided by 12 to calculate the Average Indexed Monthly Earnings (AIME). Third, the SS benefit formula is applied to AIME to produce the Primary Insurance Amount (PIA), the benefit payable at the Full Retirement Age (FRA). Finally, benefits are adjusted to produce permanently lower or higher benefits for those who claim before or after the FRA, so that the system is roughly actuarially neutral.[1] The FRA is not a static concept: it has been 65 for many years; however, beginning with people born in 1938 or later, that age gradually increases until age 67 for people born after 1959. One of the specificities of the US SS system is that retirement need not be concurrent with claiming. Indeed, fully insured individuals can claim retired worker benefits if they are at least age 62, have worked for at least 40 quarters, but ceasing work is not required, though SS takers are subjected to an earnings test: if their earnings exceed a certain amount, benefits are reduced by $1 for each $3 earned before FRA[2]. Above the FRA, the earnings test has been eliminated since 2000.

Another particular feature of this institutional structure is that spouses and survivors have the right to become entitled to special SS benefits: spousal benefits are payable when they exceed the benefit payable by reason of the woman's own earnings record (married men are also entitled to these benefits but they rarely have much in any value because married men usually have larger PIAs than their wives, and usually pre-decease them). The spousal benefit equals 50 percent of the husband's PIA when claimed at the wife's FRA. She can claim it as early as age 62, provided her

[1]Benefits are reduced by 5/9 of one percent for each month they are received prior to the FRA up to 36 months, and 5/12 of one percent thereafter. The delayed retirement credit has increased substantially over the years, from ¼ of one percent for those born between 1917 and 1924 to 2/3 of one percent for those born after 1943.

[2] This amount has been increased over years, from $9120 of annual earnings in 1998, to $37,680 today.

husband has already claimed, with a subsequent reduction. [3] In the past, if a worker delayed collecting SS, the spouse would not be able to collect spousal benefits and would not be receiving delayed retirement credits either, which reduced the value of delaying SS for many couples. Changes made under the Senior Citizens'Freedom to Work Act of 2000 allow a worker to "file and suspend" SS benefits once the FRA has been reached, allowing the spouse to begin receiving spousal benefits based on the worker's record while the worker continues to accrue delayed retirement credits. Surviving spouses of retired workers are entitled to a survivor benefit of 100% of the retired worker's benefit, which can be greater or less than his PIA, depending on the age when he first claimed benefit. The benefit can be claimed once the survivor is 60, and is subject to a reduction depending on the survivor's age when benefit begins.[4]

2. Previous literature

Social Security benefits structure has been the focus of most of the studies on the effect of SS on retirement. Some of them use aggregate data to reveal the impact of SS by examining the labor force behavior of older workers at different ages. Hurd (1990) and Ruhm(1995) find a spike in the age pattern of retirement at age 62 and show that this peak has growned over time as SS benefits have increased; besides, Burtless and Moffitt(1984) show evidence that this peak did not exist before claiming at 62 became possible. Another peak at age 65 may be the result of an unfair actuarial scheme that discourages working beyond age 65. Blau(1994) shows empirically the existence of this peak since nearly 25% of the men in the labor force on their 65th birthday retire in the next quarter in his data, which is 2.5 times more than the hazard rate of the surrounding quarters.

Another strand of the literature uses micro-data sets with SS benefits determinants or ex-post benefit levels to measure the incentives to claim via sophisticated computations of Social Security Wealth (SSW) as present discounted value of future SS entitlements. Then retirement models are estimated as functions of SSW and/or SSW accruals in the case of additional years of work. Although most of this literature concludes that SS has large effects upon retirement, these effects appear very small compared to the time trend in male retirement over the past 50 years.

As it is pointed out in Coile et al.(2001), this literature suffers from a potential weakness consisting in the endogeneity of the timing of SS benefits claiming and therefore of the benefits level. Indeed, using actual SS benefits as the key independent variable to explain SS take-up decision rather than PIAs must produce biased estimations of the impact of SS upon retirement since SS benefits are themselves a function of the timing of SS claiming. As we do not have access to restricted data, PIAs are not available in the dataset used in the following analysis, but we prefer not using SS benefits level either because of the endogeneity issue raised before.

[3] Spousal benefit is subject to a reduction of 25/36 of one percent for each month they are received prior to the FRA up to 36 months, and 5/12 of one percent thereafter. There is no delayed retirement credit.

[4] This reduction is of 0.475 percent for each month it is received prior to the wife's FRA for women born in 1939 or earlier, decreasing to 0.339 percent a month for those born in 1962 or later. It is not increased if the husband's death occurs after the wife's FRA.

Existing literature is also defincient in assuming most of the times that retirement and claiming are one unique decision. This way, there may again be some mismeasurement of the key regressor (the accrual rate for instance) since claiming can be delayed *after* retirement. In other words, if claiming is distinct from retirement as cessation of work, then it limits the impact of additional work on SSW accruals; the major impact stems from delayed claiming, and studies focusing on that of additional work may miss an essential piece of the retirement timing puzzle.

Yet, a few papers did analyze independently the SS claiming and the retirement decisions. First, Coile et al.(2001) examine the claiming behavior of single and married men who had retired before age 62, thus unable to claim benefits upon retirement. They consider the age of claiming as a purely financial decision, separate from their decision to retire. Using data from the Social Security Administration's New Beneficiary Data System (NBDS) for mid-1980 to mid-1981, they find that most men in their sample claim as soon as they become eligible, or soon thereafter. Yet a substantial minority, whose characteristics imply greater Expected Present Value of Benefits (EPVB), does delay. As such, households who would see their SSW decrease or hardly increase if they delayed-single men, those with shorter life expectancy, and married men with older wives- tend to claim quite early; on the other hand those with bigger gains at delaying-married men with younger wives and men with longer life expectancy-are more prone to delay. Household wealth is also identified to influence the age of claiming, as both tails of the wealth distribution are more likely to claim early, suggesting impatience and/or liquidity constraints at low wealth levels, and strong bequest motives at high wealth levels. Finally, even controlling for this inverse U-shape, households appear to leave significant amounts of SSW "on the table".

Second, Hurd, Smith, and Zissimopoulos(2004) analyze the SS take-up decision as an independent decision, using the first four waves of the Health and Retirement Study (HRS). The only variable they identify to explain claiming ages is subjective mortality beliefs, with little explanatory power. Again, they find that most households do not maximize their SSW, claiming too early.

Sass, Sun, and Web(2007) follow the same approach to investigate the reason why married men claim SS benefits so early. Limiting their sample to households who retired prior to becoming eligible to claim SS benefits, they find no statistically significant relationship between the age of claiming and either household wealth, or expected longevity. Caddish husbands seem not to be responsible, leaving only one explanation: educational attainment, strongly related with the age of claiming. As they control for many characteristics associated with education (income, wealth, longevity and time preference), they conclude that the effect of education entirely goes through financial awareness.

II. Empirical evidence on early claiming

1. Data

The database we use is the Health and Retirement Study, which is the first database in terms of health, retirement, and aging in the US. Initially the HRS was a longitudinal study upon individuals aged 51 to 61 at baseline (that is, in 1992), and their spouses, who could be younger or older, with new interviews of these 12,652 respondents from 7,702 households, every two years. Since 1998 the HRS have enriched of two older cohorts (born before 1924 and between 1924 and 1930), and two younger cohorts (born between 1942 and 1947, and between 1948 and 1953). Our database, provided by the RAND Corporation, is a cleaned and processed version of the HRS, which contains data about five cohorts, during eight waves separated by one or two years, from 1992 until 2006. Although the HRS is the best available database to explore retirement issues in the US as a mine of information concerning health, wealth, demographics, respondents' expectations and projects concerning retirement, *etc.*, it suffers from its discrete panel dimension. Indeed, its biennal structure makes it possible to observe transitions from one state (*e.g.* work) to another (*e.g.* retirement), but hard to study their determinants as most variables are only measured at the time of the interview, *i.e.* every even year. Thus this dataset may not be the fittest to implement a survival analysis of claiming delays, but it provides enough information to allow this kind of analysis (particularly the exact date of SS take-up), which might be the best tool to analyze SS claiming behaviors.

2. Cross-sectional analysis[5]

2.1. Sample and summary statistics

The stacked data are restricted to those who never applied for disability benefits, because they represent a separate pathway to retirement that is subject to its own particular rules, and to those who have been working at least for 10 years (40 quarters) in their working lives, so that they are eligible to SS benefits entitlement, conditional on being old enough. Besides, although every individual may be observed at several waves, only one observation –that of the wave before the 62nd birthday– will be part of this cross-sectional analysis. Hence the sample is men and women eligible to SS take-up, whether they are currently working or not. Indeed, we follow Coile et al.(2001) in considering that non-working individuals (*i.e.* early retirees) form a “cleaner” sample to study transitions to SS take-up, but once the determinants of early claiming examined for the clean and non-clean samples, it might be very interesting to investigate jointly the retirement and the claiming decisions, which will be done using a bivariate probit.

The purpose of this section is to get a clear picture of the determinants of claiming SS benefits as soon as possible, *i.e.* at age 62, for workers as well as for already retired individuals, since their claiming decisions might be driven by distinct characteristics. Then an attempt to model jointly SS take-up and retirement transitions will be made on the “non-clean sample” of working individuals, in order to check that both decisions share the same determinants, which do not impact the two transitions in the same way.

Summary statistics concerning both samples are available in the appendix. Table 1 shows that 86% of the clean sample take SS benefits upon their 62nd birthday (more precisely between the wave before and after the 62nd birthday), while roughly 50% of the non-clean sample do so. Within the clean sample (cf Table 2), key features of the pattern of claiming appear through these statistics: high educational attainment, low net wealth, having a working spouse, or a short job history seem to be associated with higher chances of delaying, while the contrary goes for having a spouse already entitled to SS benefits, or benefiting from a private pension plan covering retirees. Yet the 14% of the clean sample who do not claim early (called “postponers”) only represent 180 individuals, which is

[5] Classical panel regressions cannot be run here as SS take-up is a unique event in one’s life, so that the dependent variable would be coded « 0 » for all waves except that of SS take-up. Therefore the focus of this part will be on early claiming at age 62, with all explanatory variables will all be measured the wave before the 62nd birthday while the outcome is revealed the wave following that event.

far from being representative. The “non-clean” sample should be more representative as it groups 75% of the total sample of eligible-to-SS individuals. Within this group, other characteristics seem to drive the SS take-up decision: again, being highly educated is associated to greater chances of delaying, but this is no longer the case for the pre-cited characteristics; likewise, “having a pension plan that covers retirees”, or “being married to a working spouse”, lost their discriminating power compared to the clean sample, which reinforces the intuition that the two samples do not take their decisions based on the same criteria. Hence it seems quite rational to study the two samples separately, and to model jointly the retirement and claiming decisions for the non-clean sample, for whom both decisions might be strongly related.

These statistics may be useful to describe the samples that will be studied later, but they hardly provide any clear picture of the determinants of claiming early, or delaying. Indeed, as it has been widely explored descriptively in Panis et al.(2002), earlytakers are the mix of two very distinct groups: those gifted with “nice” characteristics on the one hand, such as having a pension plan, being covered by an (ex-)employer health insurance, enjoying an excellent health, and nice job characteristics too (high income, white-collar job, *etc*.); and those less lucky, whose health is poor, without any health insurance, more prone to be blue-collars working in hard conditions for a little pay, *etc*. In other words one may claim as early as possible just because he can afford it, and prefer taking SS benefits than drawing down his bequeathable assets, or simply because he is liquidity-constrained to do so. As our sample excludes anyone who ever claimed disability benefits, who may be at the same time the poorest, with the poorest job prospects, earlytakers in our sample may be drawn from the well-off group.

Besides, summary statistics give a hint of what kind of results are expected through observable correlations between variables, but it does not allow catching the impact of each variable independently from the others, *i.e.* ceteris paribus. For instance, the highly educated are likely to be at the same time the most wealthy and healthy, so that it is impossible to capture the impact of education upon the probability of claiming early just reading these tables. Therefore, one of the aims of the regressions that will be run thereafter is to disentangle the contribution of each explanatory variable to SS take-up decision, in order to capture for instance the impact of educational attainment when gender is controlled for.

2.2. Empirical strategy and results

In this purpose, simple probit regressions will be run, on the whole sample first, then splitting it into clean and non-clean sub-samples, to examine the contribution *ceteris paribus* of each variable on the probability to take early SS benefits between these two waves; finally retirement and claiming at age 62 will me modelled jointly though a bivariate probit.

Coile, et al.(2001) implement a theoretical examination of claiming delays, and use simulations of financial gains from delay to generate cross-sectional predictions they later test in their empirical analysis. These predictions can be compared to the results of our probit regressions.

Table 4 displays the estimates of probit regressions that are run over the whole sample, and the clean and non-clean subsamples. In order to make those estimates easier to compare with Coile’s results, the same estimates, but restricted to male only, are provided in Table 5. Many of the explanatory variables included in the equation are statistically significant when the whole sample is considered, but it seems that their explanatory power disappears within the clean sample. Indeed, high educational attainment decreases the probability of claiming SS benefits, or increases that of delaying, by roughly 10%, in the whole sample. But this marginal impact is actually of greater magnitude in the non-clean sample, and insignificant is the clean sample. Amongst men the impact of education is even greater: being highly educated raises the odds of delay by 17%. Those who went to college and above are likely to work in the best jobs in terms of material (pay and other financial advantages) and immaterial (interest of the job, responsibilities, fairness of effort-reward balance, *etc*.) conditions, but this effect only exists when individuals do have a job, therefore it loses any significant power within the clean sample.

Being single also increases the odds of delay, but in a significant way only for females, as if spousal status did not matter in SS take-up decision for men. Coile et al.(2001) 's model found that married men had greater financial incentives to delay than single men, but their predictions were at odds with their simulations. We find that single women have greater chances to delay than married women (18% more chances for the non-clean sample), probably because they are on their own, while couples are better able to self-insure, and they are working women, so they may want to delay collecting SS benefits and keep on working. Non-working women have no such significant spousal pattern, neither do men.

Being part of the two lowest quintiles of total wealth has a negative impact on early claiming only for the clean sample, which is one side of the results found in Coile et al.(2001). Indeed their simulations and regressions implied a U-shaped pattern in wealth holdings, the incentives for claiming being higher for the two tails of wealth distribution, the lower tail because of liquidity constraints, and the upper tail for bequest motives. Other specifications of our equations, including wealth holdings as a continuous variable, in logarithm, or elevated to some powers, did not highlight any U-shaped pattern.

Spouse characteristics *i.e.* if spouse is working, and if he/she already entitled to SS benefits, might have some role in claiming decision, whether it be for preference for "joint retirement" (in this case couples retire together to enjoy more leisure together) or for liquidity purpose (some couples cannot afford claiming early or stopping to work early together). Whether spouse already took SS benefits only matters for the whole and the clean samples, raising the odds of claiming by 5 and 3% respectively. Non-working individuals whose spouse already took SS benefits may wait their eligibility to SS to claim as early as possible, either because they need their SS check every month or because they prefer drawing down their public pensions than their bequeathable wealth. For working individuals, as the bivariate probit may show thereafter, claiming decision is more related to retirement decision and job characteristics. Having a working spouse raises the odds of delaying, no matter which sample is used, but more particularly for working individuals (the non-clean sample), who may wait for their spouse to retire before doing so, and claim SS benefits at the same time (since both decisions might be related for people not retired when becoming eligible).

The variable that should the most determine SS take-up is the "work" dummy, included as an explanatory variable in the whole sample, but used as discriminatory to split it into clean and non-clean samples. Its (negative) marginal effect is by far the greater in magnitude and significance:

compared to someone working the wave before his 62^{nd} birthday, who is predicted to take SS benefits with a 58% probability (the other variables being held constant at their mean values), a working individual has 87% chances of claiming early. Yet the following will show that splitting the sample into two parts and focusing on the clean sample to examine the determinants of claiming, and on the non-clean sample to examine those of retiring and claiming, yield better predictions at least for the clean sample (86% of predictions are correctly classified against 73% for the whole sample). Another remarkable feature concerning job characteristics, which are only included in the non-clean sample equation, is that contrary to the other variables, they seem to have more explanatory power for men than for women. For example, working part-time (*i.e.* less than 30 hours per week) raises the odds of claiming for both categories but it does so by 16% for women, by 26% for men. Part-time jobs are associated with a lesser attachment to the job market for they are likely to be bad-salary and insecure jobs. Hence individuals in these occupations may be willing to retire/claim early. Earnings appear through another dummy, “high income”, being 1 if the individual is in the two higher quintiles of the income distribution. But earnings are defined as last calendar year earnings, so that they may correspond to another activity than the current job. High income individuals are more prone to delay, especially women, whose probability of delaying increases by 16%. The latter prediction infers that high earnings, by loosening liquidity constraints, are incentives to delay, whereas low wealth was an incentive to delay too, which may look illogical, but the first prediction is linked with liquidity constraints, the second one with bequest motives. Last but not least, having an employer-provided pension (with current job) increases the likelihood of delay, probably because again liquidity constraints are less stringent and/or because these jobs are associated with better characteristics, and more secure.

Job history, *i.e.* the number of worked years, seems to explain a good deal of the SS take-up decision: first, those whose job history is less than 10 years are not part of the sample since they are not eligible to SS take-up; then having worked between 10 and 19 years (compared to the reference group, who has been working more than 40 years) is a strong incentive to delay, increasing the corresponding probability by 20% for the whole sample, most probably for financial reasons. Indeed SS benefits are computed on the basis of the averaged highest 35 years of salary, *i.e.* the sum of the highest 35 years of salary, which may actually be less than 35 years, averaged over these 35 years. So if someone only worked for 10 to 19 years, his Average Indexed Monthly Earnings will still be divided by 35, which will yield drastically cut SS benefits.

Being covered by some private health insurance, provided by a current or ex-employer (so that this variable also applies to the clean sample), induces an increase in wealth by reducing the risk of medical expenditures. Covered individuals are therefore more likely to finance consumption out of bequeathable wealth and delay claiming benefits. Yet, if this private health insurance does cover retirees, then the impact of health insurance on delays is reversed: for instance Table 4 (whole sample) shows that being covered by private health insurance reduces the odds of claiming by 9%, but if this insurance covers retirees, the odds are raised by roughly 14%, so that the total impact on claiming is positive (=5%). This is clearly a consequence of the possibility for individuals willing to take SS to do so before qualifying to Medicare at age 65.

Self-estimated life expectations (estimating one’s own chances of surviving upon age 75 greater than 50%) do not yield any significant impact across all the tested specifications, while they were expected to increase delays as those who believe they will live a long life have more incentives to delay in order to enjoy greater benefits at older ages. This may be explained by the exclusion from the sample of anyone who ever claimed disability benefits, which distorts the sample by over-representing healthy individuals.

The bivariate probit regression (cf Table 7) does not inform more than the previous regressions about the determinants of early claiming, since the coefficients are roughly the same as those of the simple probit over the non-clean sample. But it shows that for those individuals who continue working, SS benefit take-up decision and retirement decision share the same determinants. Indeed, the coefficient of correlation of the residuals from the two equations, whose dependent variables are *retire*, a dummy equalling 1 if the working individual ceases working (*i.e.* retires), and *liquidSS*, the same dummy as in the simple probit regressions, is 80%. A non-null "rho" (the null hypothesis is easily rejected) means that there are unobserved factors that are common to both equations, which implies in turn that both decisions are related. Amongst the variables that determine SS and retirement outcomes, some only influence the receipt of SS benefits while others only impact retirement transitions. Indeed, education, marital status, part-time work, spousal timing of retirement, and health insurance have similar effects on both probabilities, but of a lesser magnitude for the probability of retiring. Only one variable has a significant influence on both dependent variables but in opposite directions: having a private pension plan, whether it is provided by a current or ex-employer, still decreases the odds of claiming by 6%, but raises those of retiring by 11%. This may be caused by some features of employer-provided pension plans, which create strong financial incentives for the employee to leave his job at age 62, by designing pension schemes that maximize the expected present discounted value of pension benefits. Hence some individuals may leave their jobs and so retire in order to have their pension benefits maximized, but wait a few months/years to take SS benefits as pension benefits loosen their liquidity constraints.

Some other variables only impact significantly one of the two dependent variables. Being single, having high income, or a short job history, only influence the probability of claiming, while being in the left tail of wealth distribution, or having optimistic self-expectations of living past age 70, decrease the odds of retiring without affecting those of claiming. The latter result is in accordance with Delavande et al. (2006) who find no effect of subjective survival on claiming, even after using instrumental variables, but a significant impact in a bivariate probit model of retirement and claiming, except that they find a positive impact on SS delays while our results show a positive impact on retirement delays instead.

These cross-sectional analyses allow disentangling from one another the impact of observable individual characteristics on the probability of claiming SS benefits between the wave before eligibility and the wave after. But of particular interest is the impact of these characteristics on longer delays, since these are the most relevant cases for policy-making. Besides, even without focusing on the impact of chosen covariates, examining the SS claiming pattern over time, with its spikes and troughs, might be of the most relevance too.

III. Survival analysis

1. Survival analysis vs. Traditional panel data methods

The claiming decision is analyzed most naturally in a duration model framework, but few are the studies that were constructed using the survival analysis tool to examine claiming delays.

The HRS is a panel dataset, meaning that there is only information on states of the units at pre-determined survey points, but the course of the events between the survey points remains unknown. Though it contains more information than cross-sectional data, it involves well-known distortions created by the method itself. Most research in social sciences is based on observational data, meaning that manipulation of the timing of explanatory variables is not possible. For instance, to study the effect of being employed on claiming behaviour, it is impossible to observe the change from employment to non-employment at some precise wave. Thus, the longer the interval between panel waves, the more uncertainty there will be regarding the exact point in time between two waves when an individual made his transition from employment to non-employment. Even more problematic is the situation where one transits from one state to another several times between two interviews, which is all the more possible as the interval between two waves is long (2 years in the HRS case is long enough to make the causal order of explanatory and dependent variables quite ambiguous).

Besides, in this kind of observational studies, explanatory and control variables may not only change stepwise from one state to another but can often change continuously over time (*e.g.* income, wealth, age, *etc.*). Therefore the use of panel data causes an identification problem due to omitted factors whose effects are summarized in a disturbance term. As these factors are not stable over time, the disturbance term cannot be uncorrelated with the explanatory variables. Hence panel analysis is particularly sensitive to the “length of the time intervals between waves relative to the speed of the process” (Blossfeld *et al.* (2007)). A major advantage of the continuous-time observation design in survival analysis is that it makes the timing between waves irrelevant.

For all these reasons, it is worth using the survival analysis tool to examine the claiming pattern based on what is observed within the selected sample (non-parametric analysis), and to estimate the impact of some covariates upon survival until claiming, using a semi-parametric approach.

2. Survival analysis applied to SS claiming

The sample consists of every individual who took SS benefits after becoming 62, since some were able to do so at age 60. For instance, a widow or widower may be able to receive reduced widow or widower benefits as early as age 60. If the surviving spouse is disabled, benefits can begin as early as age 50. Further, a widow or widower can receive benefits at any age if she or he takes care of the deceased worker's child who is entitled to a child's benefit and younger than age 16 or disabled. A deceased worker's unmarried children who are younger than age 18 (or up to age 19 if they are attending elementary or secondary school full time) also can receive benefits. For all these exceptions to the earliest claiming age being 62, claiming before age 62 will not be considered here.

Then the variable that informs about the exact date of SS benefits take-up has many missing values, coded differently depending on the reason why the variable is missing. Individuals who claimed more than two years before entering the dataset are coded as specific missing values, and dropped from the sample. Those whose take-up date is missing for no specific reason did not entitle to SS benefits yet, therefore they are part of the survival analysis, as censored survivors.

Indeed, the process at the core of this analysis is SS benefits take-up, so that individuals live/survive from the moment they become eligible to SS early benefits, *i.e.* age 62, until they entitle to SS benefits. People can take SS benefits from age 62, so even if the event that represents birth is eligibility to SS benefits, the survival analysis will begin at age 61, in order to observe "failures" at age 62 (failure, also called "death" in this modeling, will be SS take-up). As birth and SS take-up dates are available in the HRS, the exact duration is known too, even though SS take-up occurs between two waves.

The sample at study is made of 9,338 observations, representing 6,026 subjects amongst whom 5,005 "failed" over the observation period, the others being right-censored (they do not undergo the event during their observation period). People who become older than 70 without claiming are not excluded from the sample although 70 is the maximum age to collect SS retirement benefits, because it seems hard to check if they are measurement errors or if they really do claim later than the others. These cases only represent 927 observations, and are not the focus of this analysis since delaying pattern are really crucial to understand for public policy purposes between the early retirement age and the full retirement age, or a few months after, but not so at age 70.

3. Non-parametric analysis

In this section none assumption will be made about the distribution of time to failure, nor about the form that will take the effects of the covariates. The philosophy of non-parametric approaches is to "let the data speak for itself". When no covariates is included, or when covariates are qualitative in nature, it seems natural to use Kaplan and Meier (1958) and Nelson and Aalen (1978) 's methods to estimate the probability of survival past a certain time, or to compare survival experiences for several groups discriminated by some covariates. These methods account for censoring and other characteristics of survival data.

Figure 1 plots the survivor function estimated by the Kaplan-Meier method. The Kaplan-Meier estimate of the survivor function is given by the product of one minus the number of exits divided by the number of persons at risk of exit, *i.e.* the product of one minus the exit rate at each of the survival times. Figure 2 is a "zoom" of this survivor function restricted to durations inferior than 9 years, *i.e.* until people reach age 70. Figure 3 displays the Nelson-Aalen estimator of the integrated hazard, or cumulative hazard curve, again for analysis times less than 9 years, as it will always be the case in the following. The shape of these step functions is not like a regular staircase: the height between steps varies (depending on the survivor function estimates), so too does the width of the steps (depending on the times at which failures were observed). The survivor function as well as the integrated hazard function shows a highly irregular pattern of claiming: first the "slopes" of our step functions, or the height of their steps, are higher at round ages, as people are always more prone to make transitions at round ages. Second, these higher slopes, which correspond to spikes in SS take-up, are not all of same magnitude: the two higher spikes occur at the early and normal retirement ages (62 and 65, *i.e.* 12 and 48 months of analysis time), with the second apparently greater; other spikes are observable at 24 (age 63), and 36 (age 64) months, but of lesser magnitude; there seems to be another spike of high magnitude around 60 months, which may correspond to the shift of the

NRA (normal retirement age) from age 65 to age 66 (and then age 67). After the 60th month the same pattern of higher slopes at round ages becomes hardly visible (since most of the sample took SS benefits at that point), though it still exists.

Trying to estimate the slope of the integrated hazard function at each of the observed survival time is quite tricky as it is equivalent to trying to find the slope at the corner of each of the steps. Clearly the slope cannot be well-defined, nor is any non-parametric estimate of the hazard rate. Yet one may try two methods to do so: one consists in deriving estimates of the interval hazard rate, the other in smoothing the step function (though the latter incorporates assumptions needed to “connect the dots” that may not be appropriate). Both are tested here, and displayed in figures 4 and 5, which confirms what was already clear in previous figures.

Before going further with semi-parametric estimates, it may be useful to spend a little more time on the concept of hazard rate: the hazard rate of claiming can be defined as a conditional probability. For example the hazard rate of claiming at age 63 is the probability of claiming at age 63 conditional on not having claimed until that age. Hence even though over 30% of the individuals of the sample take SS benefits within 3 months after turning 62, while hardly 14% begin collecting them in the 3 months that follow the 65th birthday, the hazard rate is much more higher at age 65 than at age 62, because the number of individuals failing is divided by the number of individuals at risk at each survival time, and the latter number strictly decreases with time.

4. Semi-parametric estimates

The empirical strategy that will be implemented in the following consists in specifying only a functional form for the influence of covariates, but leave the shape of the transition rates as unspecified as possible. In other words the model to be estimated is a semiparametric one, the Cox model, which may be written as $\mathrm{r}(\mathrm{t}) = \mathrm{h}(\mathrm{t}) \exp(\mathrm{A}(\mathrm{t})\alpha)$. The transition rate, $r(t)$, is the product of an unspecified *baseline rate*, $h(t)$, and a second term specifying the possible influences of a covariate vector $A(t)$ on the transition rate. This is a special case of so-called *proportional* transition rate models because the effects of covariates can only induce proportional shifts in the transition rate but cannot change its shape. Therefore the Cox model should only be used if this proportionality assumption is justified.

Figure 6 provides evidence in favor of the proportional-hazards assumption for almost all the covariates whose effect on claiming delays may be significant (according to previous cross-sectional analyses). These graphs plots the transformation *–ln(-ln(S(t)))* for each curve (*S(t)* being the survivor function) with *ln(t)* on the *x* axis. Except for the “pension” dummy, which equals 1 if the individual has a private pension plan from some current or ex employer, the other covariates seem to satisfy the proportional-hazards assumption. After a long period of survival, around 60 months, there seems to be a “rank inversion” for some of the tested covariates, which one may observe on the graphs when the curve representing the highest survival crosses the other curve and then stays below. Actually, this result is not surprising as the slope of the highest survival curve might become greater than that of the lowest at the point where the risk pool of the less surviving group becomes too small. For example, men survive better than women (*i.e.* they claim later) but their survival curves are perfectly parallel, until that of the men group crosses that of women, because the at-risk

pool of women has become too small since most of them already failed. This is a pure mechanical effect, which occurs after long enough delays not to bother the implementation of the Cox model.
Likewise, the highly educated survive better than their counterparts and single individuals better than married ones, in accordance with cross-sectional predictions of the previous section. These curves almost cross but remain quite parallel until the end of the period. Low wealth does not yield very distinct survival curves, but still the log-rank test rejects the null hypothesis of equality of survivor functions. Last, having high self-expectations of surviving after age 70 is associated with a higher survival.

Figure 7 applies the same method of checking the proportional hazards assumption with a highly discriminating variable which equals 1 if the individual is a self-declared "young retiree", *i.e.* declared he retired (stopped working) before age 62, 0 else. Unsurprisingly, the young retiree group survives dramatically less at every survival time until the rank inversion when delays are superior to 60 months or so. This variable has a great role in determining delays so that it has to be included in Cox regressions, but as a discriminating variable rather than an explanatory one. Indeed group effects will be modeled by estimating stratified, clustered, and shared frailty models.
First simple Cox regressions are estimated and displayed in Table 8, with the "young retiree" variable included as explanatory variable. The interpretations chosen here is that in terms of hazard ratios, which are simply exponentiated coefficients in this model[6].

Last thing to know about the Cox model before going into the tables' details is that Cox regression results are based on forming, at each failure time, the risk pool or risk set, *i.e* the collection of subjects who are at risk of failure, and then maximizing the conditional probability of failure. The times at which failures occur are not relevant in a Cox model, but the ordering of failures is (partial likelihood method). As such, when subjects are tied, *i.e.* fail at the same time, and the exact ordering of failures is unclear, the situation requires special treatment. This is particularly the case here, as many people have the same survival times (12, 48, 60 months, *etc.*). The way ties are handled depends on the reason why survival times are tied. If one believes that failures did not really occur at the same time, but the data says so because it is not continuous (for example dates are precise to the month but not to the day), then the marginal calculation method considers all possibilities (one subject failed and then the other, or conversely) and computes the sum of all these conditional probabilities. Breslow and Efron's methods are approximations of the exact-marginal calculation. If failures really occur at the same time then the partial calculation computes the right probability, but it can produce bad results when risk pools are large and there are many ties, so that we only test the other methods, in Table 8.

Whatever method for ties is used, all the variables included in the Cox regression are highly significant, and keep the same direction as in cross-sectional analyses. Women have hazard rates increased by 14% compared to men. If the survival curves were drawn for the two groups, they

[6] Indeed, for a subject with covariates $x_1, x_2, \dots, x_k$, the hazard rate under this model is $h(t|x_1, x_2, \dots, x_k) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$, so for a subject with the same covariates except that x_2 is incremented by 1, the hazard would be $h(t|x_1, x_2 + 1, \dots, x_k) = h_0(t)\exp(\beta_1 x_1 + \beta_2(x_2 + 1) + \dots + \beta_k x_k)$, and the ratio of the two hazards is thus $\exp(\beta_2)$.

would have exactly the same shape, but that of men would be above that of women. Again, high education and being single lower hazard rates of claiming. Low net wealth decreases the odds of claiming by roughly 10% compared to the 3 higher quintiles of total wealth assets. Thus, it seems that the bequest motive, which could explain why wealthy individuals collect SS benefits so soon, is a stronger incentive to claim for the wealthy than the lack of financial wealth is an incentive to delay for the less wealthy. Self-estimated probability of surviving until age 70 still does not impact delays. Having a pension from current job, which in fact corresponds to an interaction between the "work" and the "pension" dummies, decreases hazard rates by roughly 20%; being covered by some private health insurance has an even greater negative impact on failure rates, which is almost offset if this health insurance plan covers retirees. Last, declaring oneself retired before age 62 has the greatest impact on hazards since it raises them by 50%. In the following, the latter variable will be used to split the sample into two subsamples (the clean sample made of people retired before age 62, and the non-clean sample), to stratify the Cox regression, or to introduce shared frailty in the estimations. Indeed, when this variable is not included as a covariate, both groups (the clean and non-clean samples) are assumed to face the same hazard of claiming over time, while their respective risks of failing may be different, which would bias the effects of the other covariates. When it is included as a covariate, the two groups are assumed to have proportional hazards, while the hazards may be shaped differently. Therefore, fitting two models for the two groups separately might be a more proper strategy.

The first two columns of Table 9 display the results of simple Cox regressions run separately over the two sub-samples. Claiming patterns seem not to be influenced by the same variables. Females have higher hazard rates than males within the non-clean sample, while gender had no impact in the bivariate probit regression. The Cox model does not allow knowing *when* gender does impact hazard rates, but as it did not have any impact on short delays, the effect of gender may only become significant after some months. Spousal status and availability of health insurance determine claiming hazards for the non-clean sample only, in the expected directions, as it was the case in the cross-sectional analysis. Likewise, having a pension from current job decreases failure rates for the non-clean sample, and to a lesser extent for the clean sample, but only for the 12 first months of delay as those people declare they retire before age 62.

By splitting the sample in two subsamples we allowed the hazards to differ between the two groups as desired, but we also obtained separate measures of the impact of each covariate. Another solution is to stratify the analysis. Then the assumption that everyone faces the same baseline hazard, multiplied by their relative hazard, is relaxed in favor of differing baseline hazards for the two groups, but coefficients that are constrained to be the same. The hazard ratios of the stratified analysis (displayed in column 3 of Table 9) are highly significant or every covariate except life-expectation beliefs, and lie between the separate estimates for the two groups. Figure 8 shows the Cox cumulative hazard functions obtained through these regressions. The first graph corresponds to the cumulative hazards of the first two columns, i.e. when baseline and relative hazards differ between the two groups; the second graph corresponds to the third column, where only baseline hazards differ, but coefficients are the same.

Modeling group effects can also be done by clustering the analysis, in order to adjust the point estimates for the possible correlation between subjects (more precisely between subjects' failure times) within each group without making any parametric assumption about the correlation process. Estimates are showed in the fourth column of Table 9, and are quite similar to those from the stratified analysis though they produce almost confounded cumulative hazard curves for the two groups.

Last, we estimate a Cox model with shared frailty, which is simply a random-effects Cox model used to model within-group correlation: observations within a group are correlated because they share the same frailty, and the extent of the correlation is measured by θ. Here the clean and non-clean samples are assumed to differ by a frailty parameter, with the clean sample frailer than the non-clean. The hazard ratio estimates are displayed in last column of Table 9. Given the estimated frailty variance $\hat{\theta} = 0.038$ and the significance level of the likelihood ratio test of $H_0: \theta = 0$, we conclude that under this model there is significant within-group correlation. The interpretation of the hazard ratios is the same as before, except that they are conditional on frailty. For example, once intragroup correlation is accounted for, for a given level of frailty the hazard for single individuals is nine-tenth that of married individuals. As a subject frailty would have a lot to say about the hazard, it may be interesting to estimate the frailties, and produce a graph comparing the survivor curves at the lowest, mean (baseline), and highest levels of frailty. This is done in figure 9, with the high frailty corresponding to the low survival curve, *i.e.* the clean sample, and the low frailty corresponding to the high survival curve, *i.e.* the non-clean sample. If we assume now that there are 4 distinct levels of frailty: whether people are not retired at age 62, have been retired for 1 to 3 years, for 4 to 9 years, or for more than 10 years, the results of the shared frailty Cox regression show that there is again significant within-group correlation. The four distinct survivor curves are plotted in Figure 10 too. The highest curve, *i.e.* that of the least frail group, corresponds to the individuals who are not retired at age 62. The second least frail group is made of those who had already been retired for 10 years. The second frailer group had been retired for 4 to 9 years at age 62, and the frailest group is made of those newly retired (for 1 to 3 years) at age 62. Thus the longer individuals have been on retirement, the least they will delay claiming (or the longer they will "survive").

Conclusion

This paper has examined the Social Security claiming behavior of older Americans, first by casting light on the determinants of the retirement and claiming decisions at the Early Entitlement Age (*i.e.* at age 62), then by studying delays in Social Security take-up using survival analysis tools. I have two principal findings. First, Social Security take-up and retirement as cessation of work do share common observed and unobserved determinants as the bivariate probit analysis has showed. Indeed, high education, low wealth, a working spouse, and the availability of some private health insurance tend to impact negatively both transitions, while working part-time or benefiting from a private insurance covering retirees have positive impacts on these transitions. One variable though has a significant effect on the two dependent variables in opposite directions: having a pension plan provided by one's employer is associated with higher chances of retiring but lower chances of claiming Social Security.

Second, the survival analysis of the time-to-claiming has found similar results as the first cross-sectional analyses concerning short delays, but also allows investigating the determinants of longer delays. In the latter analysis the focus is on how claiming patterns differ between already "retired" individuals, *i.e.* who already stopped working, and those who are still working. Survival curves plotted for the two groups as well as the results from Cox regressions show clearly that those who ceased working "survive" far less than the others. I believe that this result deserves further research and modeling, as the retirement literature always study the incentives to take Social Security at age *a* assuming that people can and do work until that age. This paper gives the empirical motivation to build some new retirement model that would emphasize the higher incentives to claim Social Security benefits early when not earning any salary income. As the positive impact of individual inactivity on Social Security transitions has already been put forward in this study, I now intend to investigate the influence of aggregate regional unemployment on the probability to "fail" (*i.e.* to claim), for both working and non-working individuals. Peaks in aggregate unemployment have been proved to increase significantly retirement transitions (Coile and Levine, 2006), but preliminary results that I find show an opposite effect on the probability of claiming (see Table 10),which could be a first step to sharpen the analysis of Social Security claiming pattern at the core of this study.

APPENDIX

Table 1-- Summary statistics on clean and non-clean samples

Variables	Obs.	Mean	Obs.	Mean	Obs.	Mean
	clean sample		non-clean sample		Whole sample	
retires by next wave	1448	0	4130	0,25	5578	0,18
claims SS by next wave	1279	0,86	3428	0,52	4707	0,61
female	1448	0,60	4130	0,45	5579	0,49
college and above	1448	0,22	4125	0,24	5574	0,23
not married	1447	0,15	4129	0,23	5577	0,21
difference of ages with spouse	1447	1,51	4127	2,22	5574	2,03
low net wealth	1448	0,27	4130	0,37	5579	0,34
high income	1448	0,26	4130	0,54	5579	0,47
currently working	1448	0	4130	1,00	5578	0,74
works 0-29 hours per week	1448	0	4086	0,16	5534	0,12
spouse claimed SS	1193	0,42	3176	0,26	4369	0,30
spouse works	1424	0,36	4068	0,49	5492	0,45
has pension from current job	1448	0,00	4095	0,55	5543	0,41
private ins. covers retirees	1420	0,38	3637	0,42	5057	0,41
covered by current or ex-employer hlth ins.	1438	0,41	4028	0,67	5466	0,60
worked 10-19 years	1448	0,17	4130	0,06	5579	0,09
worked 20-29 years	1448	0,18	4130	0,12	5579	0,13
worked 30-39 years	1448	0,33	4130	0,24	5579	0,27
self proba of living to 75>50	1359	0,64	3672	0,64	5032	0,64
wave 1	1448	0,16	4130	0,14	5579	0,15
wave 2	1448	0,17	4130	0,14	5579	0,15
wave 3	1448	0,14	4130	0,15	5579	0,15
wave 4	1448	0,14	4130	0,15	5579	0,15
wave 5	1448	0,14	4130	0,16	5579	0,16
wave 6	1448	0,15	4130	0,15	5579	0,15

Table 2- Summary statistics-Clean sample

Variable	Obs	Mean	Obs	Mean	Obs	Mean
	postponers		earlytakers		Total Clean sample	
female	181	0,67	1098	0,59	1448	0,60
college and above	181	0,23	1098	0,19	1448	0,22
not married	181	0,15	1098	0,15	1447	0,15
difference of ages with spouse	181	1,60	1098	1,51	1447	1,51
low net wealth	181	0,35	1098	0,26	1448	0,27
high income	181	0,26	1098	0,25	1448	0,26
currently working	181	0,00	1098	0,00	1448	0,00
works 0-29 hours per week	181	0,00	1098	0,00	1448	0,00
spouse claimed SS	153	0,27	930	0,43	1193	0,42
spouse works	175	0,48	1084	0,34	1424	0,36
has pension from current job	181	0,00	1098	0,00	1448	0,00
private ins. covers retirees	175	0,26	1078	0,38	1420	0,38
covered by current or ex-employer hlth ins.	178	0,31	1091	0,41	1438	0,41
worked 10-19 years	181	0,27	1098	0,15	1448	0,17
worked 20-29 years	181	0,17	1098	0,18	1448	0,18
worked 30-39 years	181	0,33	1098	0,32	1448	0,33
self proba of living to 75>50	170	0,66	1025	0,63	1359	0,64
wave 1	181	0,21	1098	0,16	1448	0,16
wave 2	181	0,22	1098	0,17	1448	0,17
wave 3	181	0,16	1098	0,15	1448	0,14
wave 4	181	0,14	1098	0,14	1448	0,14
wave 5	181	0,13	1098	0,15	1448	0,14
wave 6	181	0,14	1098	0,14	1448	0,15

Table 3-Summary statistics-Non-clean sample

Variable	Obs.	Mean	Obs.	Mean	Obs.	Mean
	postponers		earlytakers		Total Non-Clean sample	
female	1640	0,42	1788	0,47	4130	0,45
college and above	1638	0,28	1785	0,16	4125	0,24
not married	1639	0,23	1788	0,21	4129	0,23
difference of ages with spouse	1639	2,38	1787	2,09	4127	2,22
low net wealth	1640	0,37	1788	0,38	4130	0,37
high income	1640	0,62	1788	0,43	4130	0,54
currently working	1640	1,00	1788	1,00	4130	1,00
works 0-29 hours per week	1632	0,08	1757	0,25	4086	0,16
spouse claimed SS	1331	0,18	1456	0,30	3176	0,26
spouse works	1617	0,51	1761	0,47	4068	0,49
has pension from current job	1630	0,63	1770	0,46	4095	0,55
private ins. covers retirees	1420	0,42	1613	0,41	3637	0,42
covered by current or ex-employer hlth ins.	1604	0,73	1731	0,59	4028	0,67
worked 10-19 years	1640	0,06	1788	0,06	4130	0,06
worked 20-29 years	1640	0,11	1788	0,13	4130	0,12
worked 30-39 years	1640	0,25	1788	0,22	4130	0,24
self proba of living to 75>50	1461	0,65	1582	0,63	3672	0,64
wave 1	1640	0,19	1788	0,14	4130	0,14
wave 2	1640	0,16	1788	0,16	4130	0,14
wave 3	1640	0,18	1788	0,15	4130	0,15
wave 4	1640	0,20	1788	0,16	4130	0,15
wave 5	1640	0,20	1788	0,17	4130	0,16
wave 6	1640	0,08	1788	0,14	4130	0,15

Table 4 - Probit results

Dependent variable: LiquidSS=1 if takes SS benefits between the wave before and after 62nd birthday

	whole sample	clean sample	non-clean sample
female	-0.0300	-0.0177	-0.0207
	(0.0235)	(0.0176)	(0.0295)
college and above	-0.0941***	-0.0240	-0.118***
	(0.0245)	(0.0203)	(0.0293)
not married	-0.0833**	-0.0144	-0.107**
	(0.0290)	(0.0228)	(0.0356)
difference of ages with spouse	0.00594	-0.00269	0.00883
	(0.00394)	(0.00329)	(0.00481)
low net wealth	-0.0353	-0.0474*	-0.0109
	(0.0205)	(0.0191)	(0.0254)
high income	-0.114***		-0.117***
	(0.0224)		(0.0271)
currently working	-0.295***		
	(0.0202)		
works 0-29 hours per week	0.219***		0.262***
	(0.0202)		(0.0283)
spouse claimed SS	0.0544*	0.0326*	0.0425
	(0.0242)	(0.0154)	(0.0325)
spouse works	-0.133***	-0.0659**	-0.138***
	(0.0227)	(0.0202)	(0.0287)
has pension from current job	-0.0676**		-0.0626*
	(0.0245)		(0.0276)
private ins. covers retirees	0.139***	0.0483	0.150***
	(0.0247)	(0.0318)	(0.0294)
covered by current or ex-employer hlth ins.	-0.0955***	-0.0387	-0.127***
	(0.0283)	(0.0418)	(0.0337)
worked 10-19 years	-0.193***	-0.115**	-0.135**
	(0.0392)	(0.0382)	(0.0519)
worked 20-29 years	-0.0199	-0.0148	-0.0125
	(0.0304)	(0.0258)	(0.0379)
worked 30-39 years	-0.0469*	-0.0553*	-0.0302
	(0.0237)	(0.0227)	(0.0294)
self proba of living to 75>50	-0.0109	-0.0115	-0.00702
	(0.0186)	(0.0134)	(0.0237)
Observations	3183	984	2199
Pseudo R-squared	0.200	0.118	0.135
Observed P(liquidSS)	0,6315	0,8618	0,5284
Predicted P(liquidSS)	0,685	0,9383	0,5525
% correctly classified	73,14	86,08	66,71

Marginal effects; Standard errors in parentheses

* p<0.05 ** p<0.01 *** p<0.001

Table 5-Probit results for men

	whole sample	clean sample	non-clean sample
college and above	-0.142***	-0.0115	-0.174***
	(0.0362)	(0.0317)	(0.0422)
not married	-0.0137	0.0369	-0.0356
	(0.0395)	(0.0279)	(0.0489)
difference of ages with spouse	0.00546	-0.00456	0.00943
	(0.00443)	(0.00392)	(0.00547)
low net wealth	-0.00669	-0.0604	0.0183
	(0.0313)	(0.0443)	(0.0374)
high income	-0.0547		-0.0579
	(0.0325)		(0.0415)
currently working	-0.332***		
	(0.0260)		
works 0-29 hours per week	0.266***		0.349***
	(0.0251)		(0.0365)
spouse claimed SS	0.0248	-0.0316	0.0473
	(0.0435)	(0.0494)	(0.0545)
spouse works	-0.121***	-0.0677	-0.125***
	(0.0310)	(0.0395)	(0.0372)
has pension from current job	-0.0474		-0.0434
	(0.0356)		(0.0410)
private ins. covers retirees	0.147***	0.0942	0.164***
	(0.0378)	(0.0830)	(0.0432)
covered by current or ex-employer I	-0.0930*	-0.0546	-0.123*
	(0.0408)	(0.0541)	(0.0498)
worked 10-19 years	-0.362*	-0.372	-0.300
	(0.149)	(0.238)	(0.188)
worked 20-29 years	-0.0868	-0.0707	-0.0725
	(0.0811)	(0.0841)	(0.0999)
worked 30-39 years	-0.0460	-0.0597	-0.0348
	(0.0379)	(0.0336)	(0.0480)
self proba of living to 75>50	-0.0159	-0.00684	-0.0157
	(0.0277)	(0.0258)	(0.0342)
Observations	1347	332	1015
Pseudo R-squared	0.206	0.119	0.133

Marginal effects; Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001

Table 6-Probit results for women

	whole sample	clean sample	non-clean sample
college and above	-0.0459	-0.0208	-0.0612
	(0.0329)	(0.0247)	(0.0425)
not married	-0.136**	-0.0429	-0.185**
	(0.0439)	(0.0353)	(0.0579)
difference of ages with spouse	0.0130	0.00305	0.0159
	(0.00909)	(0.00856)	(0.0112)
low net wealth	-0.0397	-0.0341	-0.0232
	(0.0264)	(0.0202)	(0.0349)
high income	-0.158***		-0.164***
	(0.0312)		(0.0362)
currently working	-0.232***		
	(0.0286)		
works 0-29 hours per week	0.163***		0.191***
	(0.0278)		(0.0382)
spouse claimed SS	0.0429	0.0381	0.00360
	(0.0314)	(0.0195)	(0.0456)
spouse works	-0.153***	-0.0648**	-0.179***
	(0.0336)	(0.0234)	(0.0474)
has pension from current job	-0.0797*		-0.0741*
	(0.0329)		(0.0374)
private ins. covers retirees	0.113***	0.0344	0.123**
	(0.0318)	(0.0335)	(0.0402)
covered by current or ex-employer hlth ins.	-0.0974*	-0.0341	-0.128**
	(0.0381)	(0.0551)	(0.0457)
worked 10-19 years	-0.183***	-0.109**	-0.131*
	(0.0419)	(0.0406)	(0.0557)
worked 20-29 years	-0.0168	-0.0154	-0.0111
	(0.0333)	(0.0298)	(0.0423)
worked 30-39 years	-0.0441	-0.0576	-0.0237
	(0.0304)	(0.0321)	(0.0380)
self proba of living to 75>50	-0.000109	-0.0118	0.0112
	(0.0244)	(0.0152)	(0.0329)
Observations	1836	652	1184
Pseudo R-squared	0.207	0.140	0.153

Marginal effects; Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 7- Bivariate probit-Marginal effects

Sample: working individuals, the wave before 62nd birthday

	P(retire=1)	P(liquidSS=1)
female	-0.0339	-0.0238
	(0.0256)	(0.0293)
college and above	-0.100***	-0.115***
	(0.0230)	(0.0291)
not married	-0.0397	-0.107**
	(0.0294)	(0.0350)
difference of ages with spouse	0.00127	0.00935
	(0.00406)	(0.00481)
low net wealth	-0.0670**	-0.00819
	(0.0214)	(0.0252)
high income	0.00968	-0.115***
	(0.0242)	(0.0270)
works 0-29 hours per week	0.0859**	0.245***
	(0.0292)	(0.0281)
spouse claimed SS	0.00643	0.0439
	(0.0279)	(0.0321)
spouse works	-0.0937***	-0.135***
	(0.0241)	(0.0282)
has pension from current job	0.112***	-0.0594*
	(0.0237)	(0.0273)
private ins. covers retirees	0.123***	0.147***
	(0.0277)	(0.0297)
covered by current or ex-employer hlth ins.	-0.101**	-0.119***
	(0.0320)	(0.0337)
worked 10-19 years	0.00979	-0.124*
	(0.0446)	(0.0508)
worked 20-29 years	0.0614	0.0000659
	(0.0341)	(0.0375)
worked 30-39 years	0.00259	-0.0269
	(0.0255)	(0.0291)
self proba of living to 75>50	-0.0412*	-0.00514
	(0.0207)	(0.0235)
Observations	2199	
rho	0.794	
chi2(1): test of rho=0	588.2	

Marginal effects; Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 8- Cox regressions-Total sample

Method for ties	Breslow	marginal	efron
female	1.136***	1.129**	1.141***
	(0.0390)	(0.0511)	(0.0392)
college and above	0.666***	0.685***	0.655***
	(0.0281)	(0.0366)	(0.0276)
not married	0.892**	0.904	0.888**
	(0.0373)	(0.0491)	(0.0371)
low net wealth	0.902**	0.927	0.896**
	(0.0330)	(0.0442)	(0.0328)
self proba of living to 75>50	0.942	0.932	0.943
	(0.0321)	(0.0418)	(0.0321)
has pension from current job	0.810***	1.041	0.802***
	(0.0344)	(0.0560)	(0.0340)
covered by current or ex-employer hlth ins.	0.765***	0.761***	0.759***
	(0.0431)	(0.0536)	(0.0428)
private ins. covers retirees	1.224***	1.226**	1.227***
	(0.0655)	(0.0790)	(0.0657)
retired before 62	1.489***	1.013	1.533***
	(0.0609)	(0.0645)	(0.0627)
Observations		6233	
Number of subjects		4833	
Number of failures		3793	
chi2	420.7	93.46	464.9

Exponentiated coefficients; Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 9- Cox regressions-by (young retiree)

Sub-sample	Clean	Non-clean	Stratified	Clustered	Frailty
female	0.999	1.164***	1.127***	1.131**	1.136***
	(0.0669)	(0.0472)	(0.0388)	(0.0538)	(0.0390)
college and above	0.684***	0.671***	0.673***	0.688***	0.667***
	(0.0522)	(0.0340)	(0.0283)	(0.0129)	(0.0281)
not married	0.851	0.900*	0.895**	0.899***	0.892**
	(0.0732)	(0.0435)	(0.0375)	(0.00643)	(0.0373)
low net wealth	0.771**	0.978	0.927*	0.881	0.902**
	(0.0612)	(0.0408)	(0.0340)	(0.114)	(0.0330)
self proba of living to 75>50	0.984	0.926	0.943	0.939**	0.942
	(0.0653)	(0.0367)	(0.0321)	(0.0196)	(0.0321)
has pension from current job	0.688*	0.797***	0.797***	0.704**	0.808***
	(0.101)	(0.0375)	(0.0339)	(0.0813)	(0.0342)
covered by current or ex-employer hlth ins.	0.905	0.787***	0.783***	0.761***	0.765***
	(0.176)	(0.0484)	(0.0443)	(0.0440)	(0.0431)
private ins. covers retirees	0.942	1.270***	1.228***	1.311***	1.226***
	(0.183)	(0.0715)	(0.0656)	(0.00971)	(0.0656)
Observations	1396	4837	6233	6233	6233
Number of subjects	1245	3675	4833	4833	4833
Number of failures	1000	2793	3793	3793	3793
Wald chi2	57.02	202.6	241.4		243.4
Theta					0.0382
LR test of theta=0					80.54

Breslow method for ties
Exponentiated coefficients; Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 10- Cox regressions on total sample with unemployment rate

	Total Sample			non-working		working	
female	1.085*	1.063	1.063	1.209***	1.180***	0.893	0.880*
	(0.0382)	(0.0375)	(0.0375)	(0.0527)	(0.0516)	(0.0555)	(0.0550)
college and above	0.683***	0.687***	0.687***	0.715***	0.721***	0.651***	0.651***
	(0.0288)	(0.0289)	(0.0290)	(0.0363)	(0.0366)	(0.0495)	(0.0496)
not married	0.914*	0.901*	0.901*	0.896*	0.887*	0.894	0.878
	(0.0383)	(0.0378)	(0.0378)	(0.0468)	(0.0463)	(0.0660)	(0.0649)
low net wealth	0.873***	0.877***	0.876***	1.007	1.005	0.747***	0.756***
	(0.0319)	(0.0321)	(0.0321)	(0.0456)	(0.0456)	(0.0470)	(0.0475)
self proba of living to 75>50	0.946	0.955	0.953	0.880**	0.881**	0.972	0.988
	(0.0322)	(0.0326)	(0.0326)	(0.0377)	(0.0378)	(0.0550)	(0.0561)
has pension from current job	0.791***	0.789***	0.789***	0.786***	0.779***		
	(0.0358)	(0.0357)	(0.0357)	(0.0377)	(0.0373)		
covered by current or ex-employer hlth ins.	0.784***	0.779***	0.779***	0.754***	0.759***	1.073	1.057
	(0.0445)	(0.0442)	(0.0442)	(0.0479)	(0.0483)	(0.173)	(0.171)
private ins. covers retirees	1.262***	1.272***	1.273***	1.303***	1.315***	0.854	0.863
	(0.0674)	(0.0680)	(0.0681)	(0.0740)	(0.0748)	(0.140)	(0.142)
currently working	**0.789*** **	**0.783*** **	**0.883**				
	(0.0330)	(0.0328)	(0.136)				
regional unemployment rate		**0.920*** **	**0.934** **		**0.912*** **		**0.932** **
		(0.0121)	(0.0207)		(0.0149)		(0.0207)
regional UR*currently working			**0.978**				
			(0.0268)				
Observations	6233	6229	6229	4282	4279	1951	1950
N_sub	4833	4832	4832	3238	3237	1692	1692
N_fail	3793	3792	3792	2412	2412	1381	1380
chi2	362.1	400.3	401.0	173.4	204.0	58.63	68.86

Exponentiated coefficients; Standard errors in parentheses
* p<0.05 ** p<0.01 ***p<0.001

Figure 1- Survivor function: Kaplan-Meier estimate

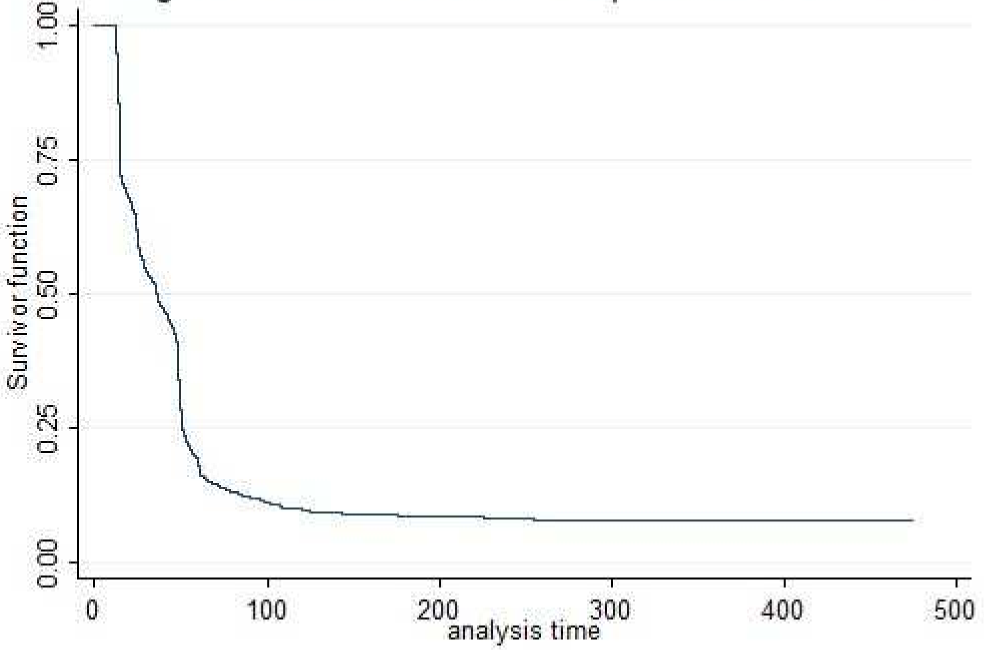


Figure 2- Survivor function

Restricted to survival times inferior to 9 years.

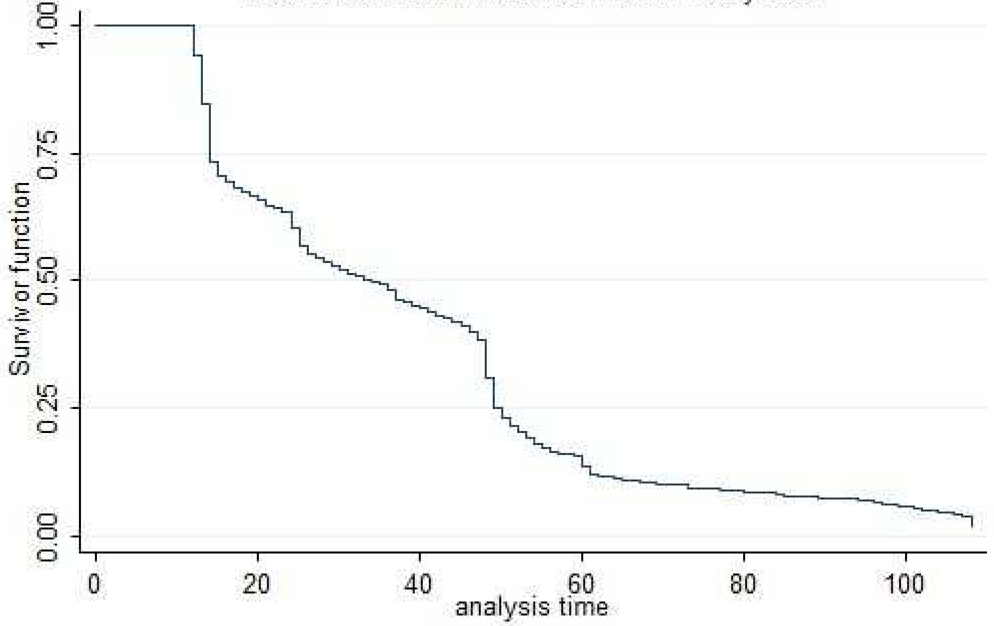


Figure 3- Cumulative hazard function

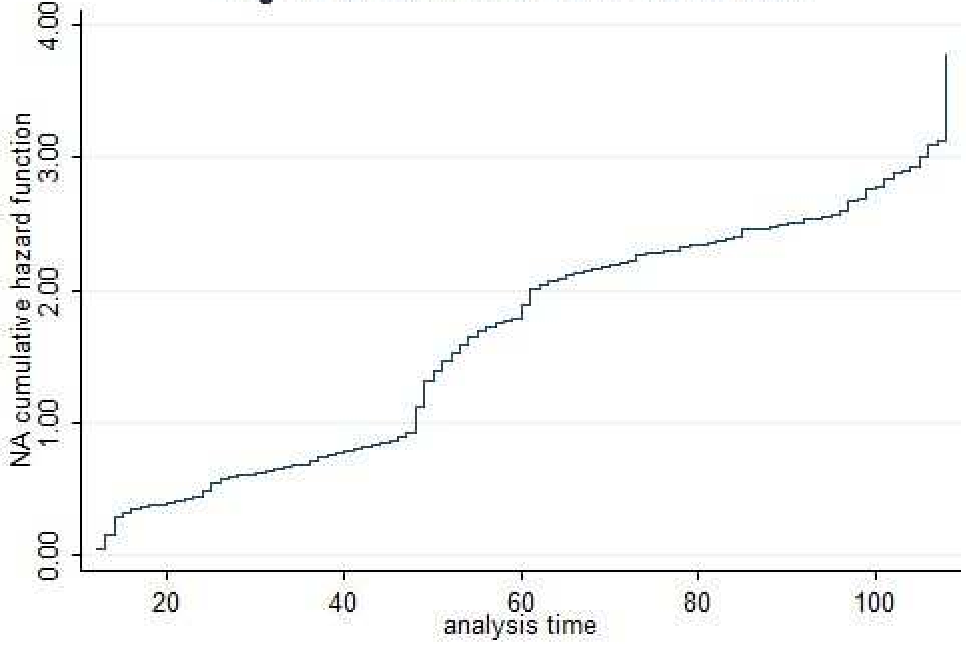
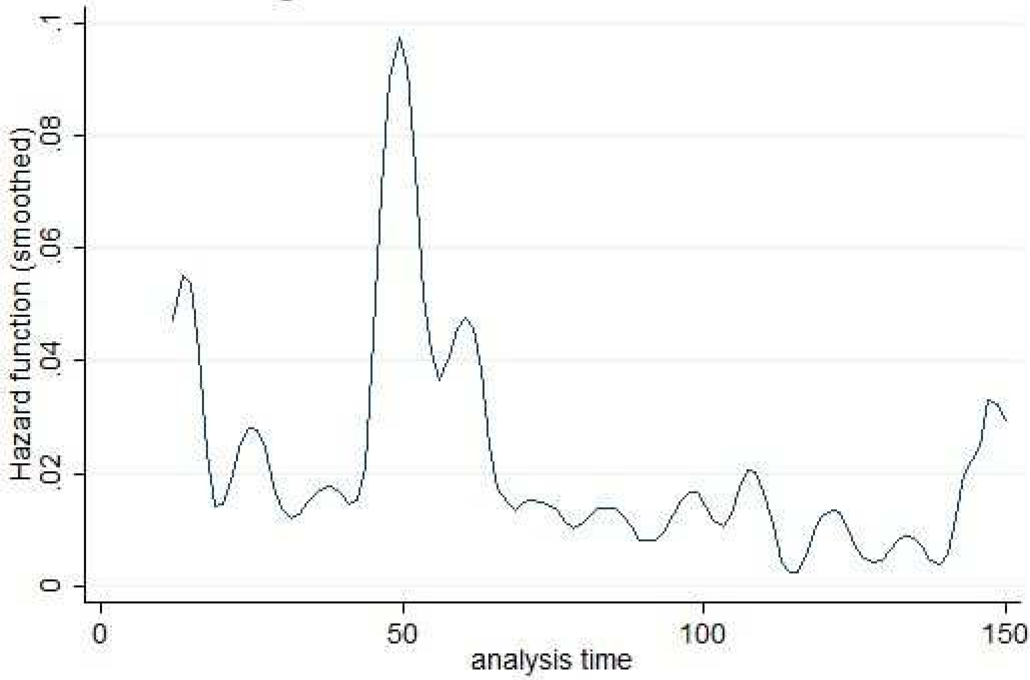


Figure 4- Smoothed hazard estimate



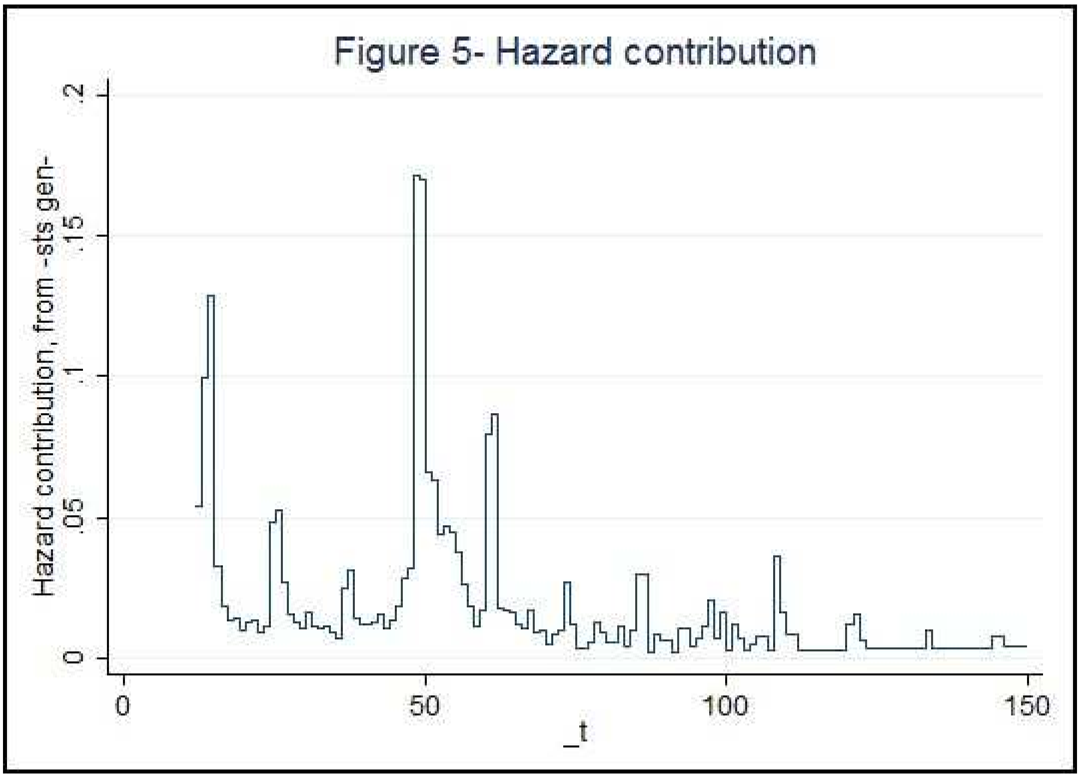
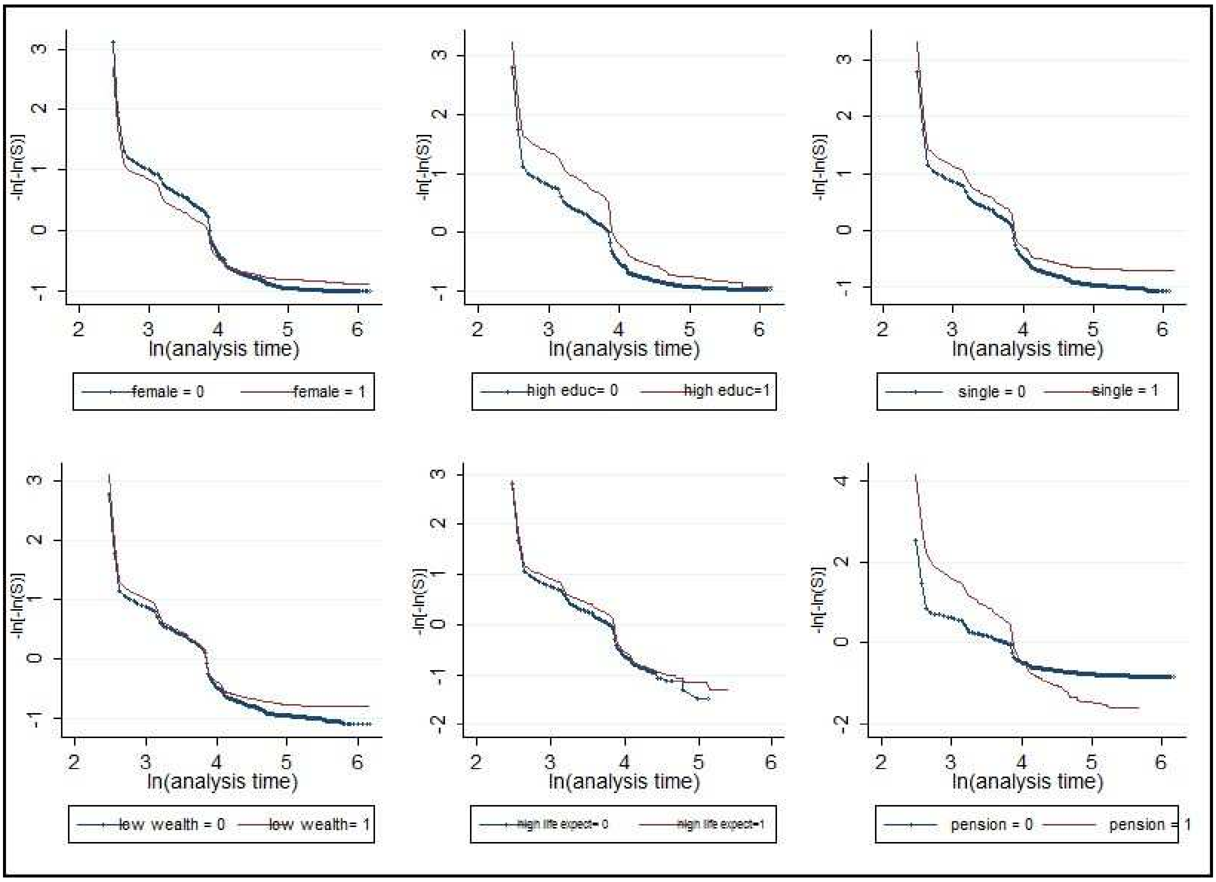


Figure 6: Test of Proportional Hazard Assumption



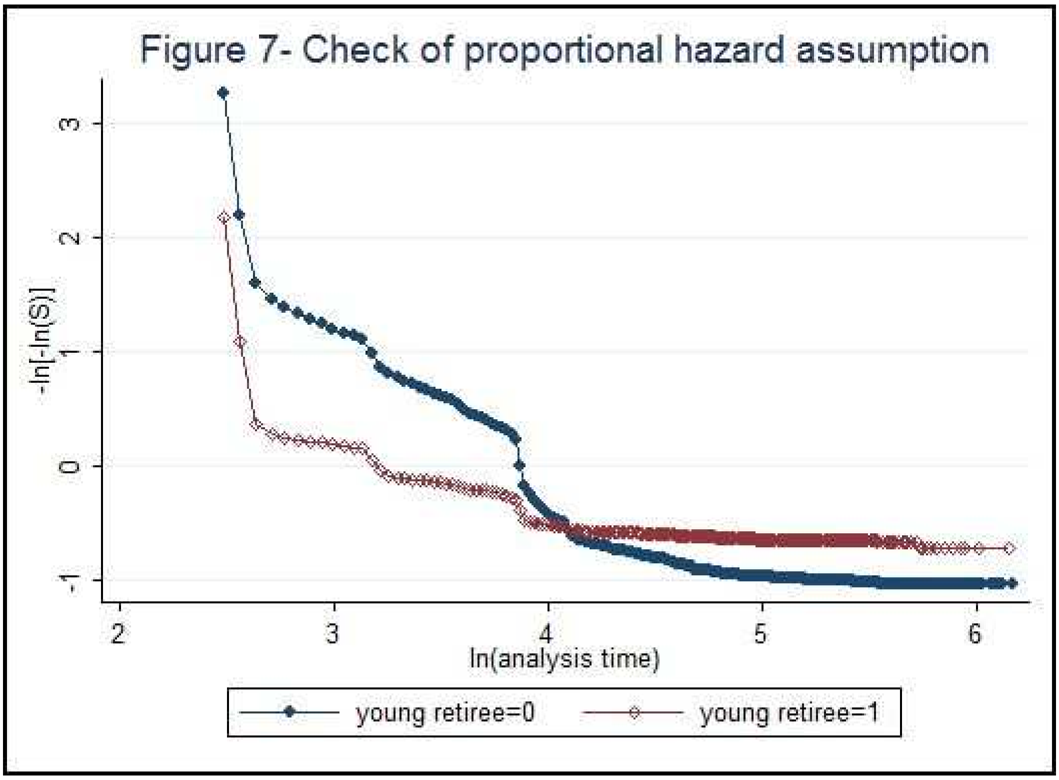


Figure 7- Check of proportional hazard assumption
-ln[-ln(S)]
3
2
1
0
-1
2
3
4
5
6
ln(analysis time)
young retiree=0
young retiree=1

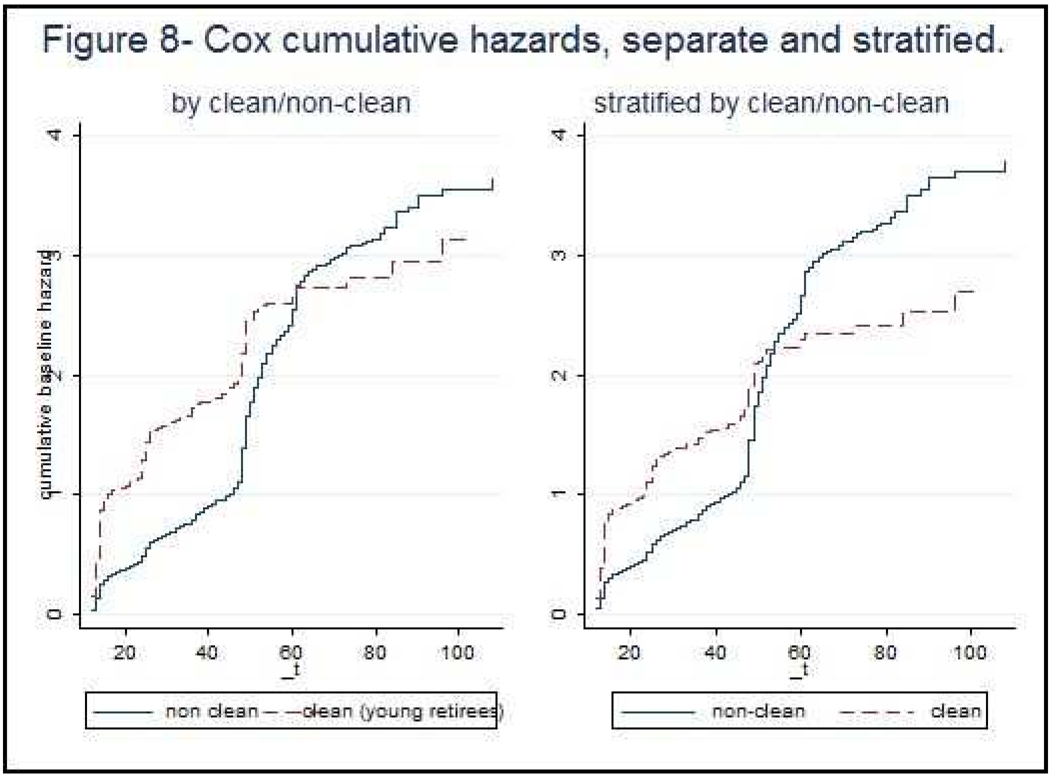


Figure 8- Cox cumulative hazards, separate and stratified.
by clean/non-clean
stratified by clean/non-clean
cumulative baseline hazard
0
1
2
3
4
20
40
60
80
100
_t
non clean
clean (young retirees)
non-clean
clean

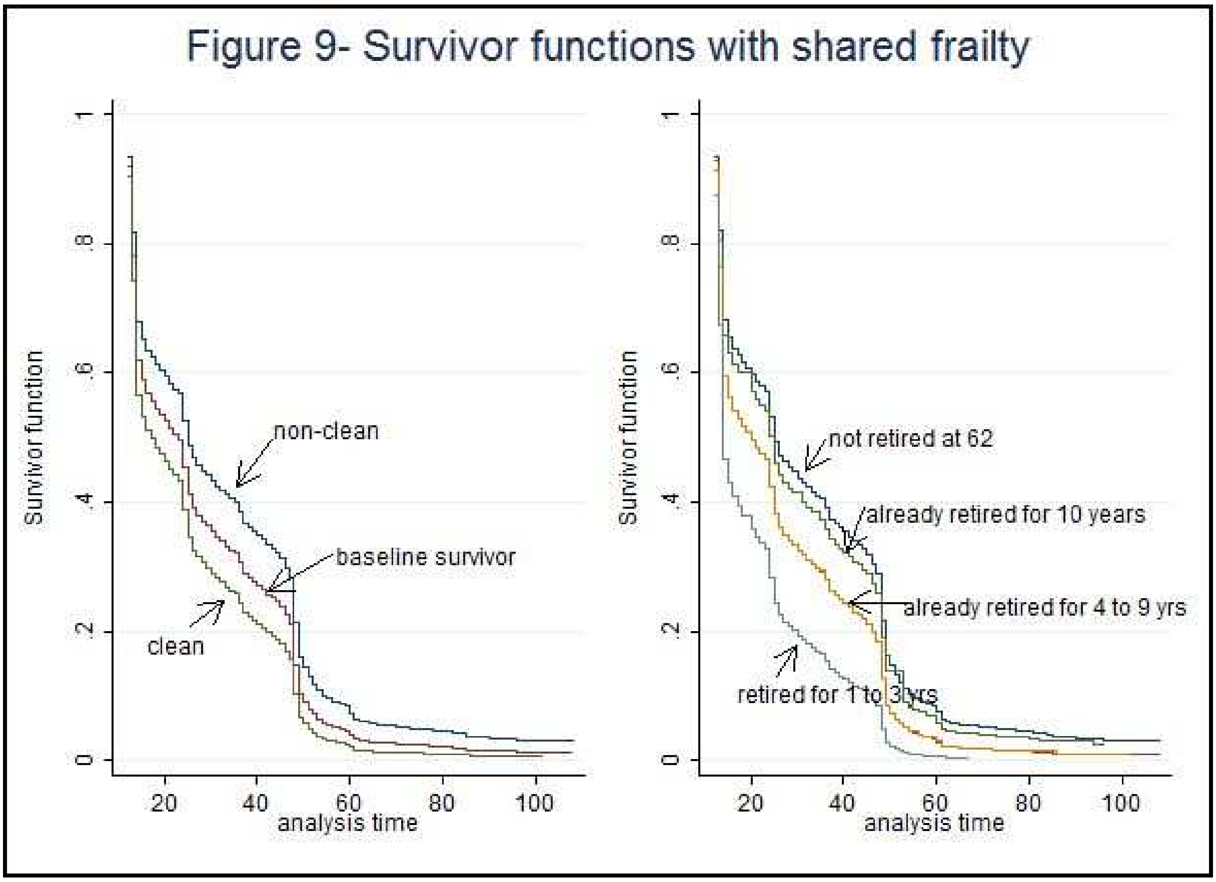


Figure 9- Survivor functions with shared frailty
Survivor function
0
.2
.4
.6
.8
1
20
40
60
80
100
analysis time
non-clean
baseline survivor
clean
not retired at 62
already retired for 10 years
already retired for 4 to 9 yrs
retired for 1 to 3 yrs

BIBLIOGRAPHY

Burtless, G., Moffitt, R., 1984. *The Effect of Social Security Benefits on the Labor Supply of the Aged.* In: Aaron, H., Burtless, G. (Eds.), Retirement and Economic Behavior. Brookings Institution, Washington, DC, 135-175.

Blau, D.M., 1994. *Labor Force Dynamics of Older Men*. Econometrica 62, 117-156.

Blossfeld H.-P., Golsch K., Rohwer G., 2007. *Event History Analysis Using Stata*. Edited by: Lawrence Erlbaum Associates.

Coile C., Diamond P., Gruber J., Jousten A., 2001. *Delays in Claiming Social Security Benefits*. Journal of Public Economics 84, 357-385.

Coile C., Levine P., 2007. *Labor Market Shocks and Retirement: Do Government Programs Matter*?. NBER Working Paper No W12559.

Delavande A., Perry M., Willis R.J., 2006. *Probabilistic thinking and Social Security Claiming*. Working Paper 2006-129. Chestnut Hill, MA: Center for Retirement Research at Boston College.

Hurd,M., Smith J. P., Zissimopoulos J.M, 2004. *The Effects of Subjective Survival on Retirement and Social Security Claiming*. Journal of Applied Econometrics 19, 761-775.

Hurd, Michael, 1990. *Research on the Elderly: Economic status, Retirement, and Consumption and Saving*. Journal of Economic Literature 28, 565-637.

Mahaney J., Carlson P., 2006. *Innovative Strategies to help Maximize Social Security Benefits*. Iselin, NJ: Prudential Financial.

Panis C., Hurd M., Loughran D., Zissimopoulos J., Haider S., StClair P., 2002. *The effects of Changing Social Security Administration's Early Entitlement Age and Normal Retirement Age*. RAND Working Paper.

Ruhm C., 1995. *Secular Changes in the Work and Retirement Patterns of Older Men*. Journal of Human Resources 30, 362-385.

Sass S., Sun W., Webb A., 2007. *Why Do Married Men Claim Social Security Benefits So Early? Ignorance, Caddishness, or Something Else*. Working Paper 2007-17. Chestnut Hill, MA: Center for Retirement Research at Boston College.