CENTER FOR APPLIED ECONOMICS AND POLICY RESEARCH

CAEPR Working Paper
#2008-016

Subjective Health Expectations

Juergen Jung
Indiana University Bloomington

June 13, 2008

Subjective Health Expectations

Juergen Jung*

Indiana University - Bloomington

9th June 2008

Abstract

In this paper I derive subjective health expectations using the RAND-HRS data. These expectations can be used in the estimation of structural life-cycle models. I use a Bayesian updating mechanism in order to correct for focal point responses and reporting errors of the orignial health expectations variable. In addition, I test the quality of the health expectations measure and describe its correlation with various health indicators and other individual characteristics. I find that subjective health expectations do contain additional information that is not incorporated in subjective mortality expectations and that the rational expectations assumption cannot be rejected for subjective health expectations.

JEL Classification: I10, D84, C11, C23

Keywords: Subjective Health Expectations, Rational Health Expectations, Work Limiting Health Problems, Bayesian Updating of Expectations.

*I would like to thank Gerhard Glomm, Kim Huynh, Michael Kaganovich, Wayne R. Gayle, Rusty Tchernis and Pravin Trivedi for many helpful comments. I am also grateful to Li Gan for making Matlab code available to me. Corresponding Author: Juergen Jung, Department of Economics, Wylie Hall, Indiana University, Bloomington, IN, 47405, phone: (812) 855-7715, e-mail: juejung@indiana.edu

1 Introduction

In economics it is a common practice to base dynamic models on agents with rational expectations. In such models, decision makers form beliefs about future income, health, and life-expectation using objective probability distributions. More recently, this practice has been criticized as economists started to directly measure subjective expectations and to document the consequences of deviating from the rational expectations hypothesis. Manski (2004) provides an overview of this literature.

In this paper I analyze subjective health expectations as reported in the Health Retirement Survey, a panel data set that covers the years $1992-2004$. This data provides information about individuals' expectations about future work limiting health problems. My results can be summarized as follows. Standard health indicators are strongly correlated with work limiting health problems. Women are less likely to develop work limiting health problems. Subjective health expectations seem to consistently predict health outcomes. The rational expectations hypothesis about subjective health expectations cannot be rejected. Younger cohorts are more pessimistic about their future health than older cohorts. There is weak evidence that individuals can learn about their health. Finally, I construct subjective health expectations following the procedure in Gan, Hurd and McFadden (2003).

An often cited problem with health survey data is that people tend to be overly optimistic about how their health compares to the average health of their age cohort. This problem becomes more pronounced with increasing age, so that one can plausibly argue that individuals' subjective expectations about future health problems might be too optimistic.[1] However, if this is the case and agents base their decisions on their subjective beliefs then using observed outcome probabilities as proxy variables for subjective expectations will introduce a bias into health uncertainty models.[2]

The health expectation that I investigate in this paper is the expectation to develop a work limiting health problem within the next ten years. In order to investigate this particular expectation, I first analyze what constitutes a work limiting health problem. Next, I analyze how well subjective health expectations predict health outcomes and finally I derive, what I call, subjective health expectations curves. I use seven waves of the RAND-HRS data for this purpose.

Subjective health expectations curves are useful since they provide additional information on agent expectations in a systematic way that can be used in the estimation of structural life cycle models. I use the method in Gan, Hurd and McFadden (2003) and extend its application to derive subjective health expectations curves. Gan, Hurd and McFadden (2003) derive subjective survival expectations based on subjective mortality expectations and data from U.S. mortality tables. Since I do not have a "health table" pendant to mortality tables, I

[1] See Eriksson, Unden and Elofsson (2001). Ludwig and Zimper (2007) find similar results for subjective mortality expectations. Younger cohorts are shown to underestimate their survival probabilities whereas older cohorts tend to overestimate their survival probabilities. Elder (2007) reports that older cohorts fail to revise their mortality expectation in the presence of increased longevity.

[2] Gan et al. (2004) use subjective mortality expectations curves and show that these curves perform better than the mortality rates that can be found in life-tables in two ways. First, they use more available information and therefore reduce the bias when estimating structural life-cycle models. Second, they perform better in terms of in-sample forecasts.

Another example that shows how important subjective survival expectations are is Bloom et al. (2006). They estimate the effects of subjective life expectations on the retirement and wealth accumulation decisions of U.S. households. They find that an increase in the perceived (subjective) probability of survival, increases the wealth of the household but does not affect the retirement decision.

construct "health tables" from the RAND-HRS data and later update these tables with information contained in the subjective health expectations.

An alternative method to estimating subjective expectations has been proposed by Perozek (2005). She fits Weibull and Gompertz distributions to subjective mortality expectations and finds that information contained in subjective mortality expectations is a predictor for adjustments made in (objective) mortality tables a decade later. In this sense, subjective expectations do contain important information that is not otherwise observed.[3] Ludwig and Zimper (2007) develop a model of Bayesian learning which combines rational learning with the possibility that the interpretation of new information is prone to psychological attitudes like initial biases and ambiguity. They conclude that rational Bayesian learning is rejected by the data. Finally, Perry (2005) uses a linearity assumption on conditional survival probabilities to construct subjective survival probabilities using the HRS. His treatment of focal point responses is ad hoc and results in lower explanatory power. He therefore drops focal point respondents out of his sample.

2 Method

In order to update subjective health expectations with information contained in the health tables we will need two variables in our data: the realization of a health-event and a variable that measures the subjective expectation that each individual has about this event.

2.1 Work Limiting Health Problems

The health variable that I use for this analysis is "work limiting health problems", known as $r.htlhlm$ in the RAND-HRS data. This variable indicates whether an impairment or health problem exists and limits the kind or amount of paid work the respondent is able to perform. Unfortunately the Health and Retirement Study (HRS) only asks whether respondents have a work limiting health problem. It is not clear how this variable is actually defined. Since I am interested in what constitutes a work limiting health problem, I qualify "work limiting health problems" by running regressions on various health indicators as well as on demographic and income variables.

The RAND-HRS data set is very rich in detailed questions about the health status of its respondents, so that regressions of this form will give an indication of which health problems are more likely to constitute work limiting health problems.[4]

In addition, the RAND-HRS data contain numerous questions about household expectations. In particular, the survey asks respondents about their subjective probability to having a work impairing health problem

[3]Her method of fitting Weibull and Gompertz distributions to subjective mortality expectations requires two observational points of subjective mortality to identify the model. The HRS can be used because in includes two measures on subjective mortality, P_{75} the probability to live to age 75 and P_{85} the probability to live to age 85. In my case however I only have one expectational measure for future health problems so that I am not able to identify the two parameters of the Weibull distribution using Perozek's method. Obviously, one could create a second data point by claiming that the probability of having a work limiting health problem at a very high age is close to one. We do not follow this method in this paper.

[4]An alternative specification includes a regression of income on work limiting health problems. This will give an indication of the income loss incurred after a work limiting health problem occurred, which will be important in setting up structural life cycle models that try to explain the effects of health and health expectations on consumption, savings and "life-style" behavior. This question, however, is beyond the analysis in this paper.

within the next 10 years. The variable is denoted $r.worklm$. Further tests on panel type regressions will explain whether expectations about work limiting health problems are accurate and formed rationally.

2.2 Health Outcome Tables and Subjective Health Expectations

Next I construct health tables that record the fraction of the population per age group acquiring a work limiting health problem within a certain period. This information is summarized in so called 'health-problem hazard rates', which is a measure of the risk that an individual runs in developing a health problem at a certain age. These tables can be used in a similar way to Gan, Hurd and McFadden (2003)'s use of U.S. decennial life tables.[5] I am then able to derive adjusted subjective health curves or adjusted subjective health hazard rates that condition on age, sex and subjective health expectations. This will automatically capture a lot of detail about an individual without having to condition on other household characteristics (e.g. income group, smoker versus non-smoker, education, etc.).

2.3 Focal Point Responses

One problem with constructing individual health probabilities are focal point responses. In wave 1, 18.20% of respondents indicate a zero probability of acquiring a work limiting health problem within the next 10 years, whereas 4.73% think that they will have a work limiting health problem with probability one. In wave 2 the respective numbers are 17.16% and 4.58% and in wave 3 they are 17.61% and 5.15%. The third focal response is at the probability of one-half. Roughly 30% over all waves respond that they expect work limiting health problems with probability one-half. I report the distribution of subjective health expectations in the histograms in figures 2 and 3 for health expectations over all six waves according to gender.

Expectations of zero and one are not very sensible. Perry (2005) finds that individuals answering with focal responses of zero and one on an expected mortality question are on average less educated, hold fewer assets and have lower income than the rest of the sample. Respondents reporting a 50% chance of surviving up to a target age look essentially the same as the rest of the sample. He therefore suggests that answers of zero and one may be more a sign of poor understanding of the question than of optimism or pessimism. I report similar summary statistics grouped by subjective health expectations of the age group $40-60$ in table 8. Respondents who report a 100% chance of developing work limiting health problems have on average lower income, asset holdings and education. All other focal respondents (0% and 50% probability of developing work limiting health problems) are similar to the rest of the sample.

The Bayesian updating model developed in Gan, Hurd and McFadden (2003) corrects for focal responses. In this model it is assumed that the prior survival probability distribution at a future point in time is a truncated normal between zero and one. The conditional density of the observed survival probability is assumed to be a censored normal between zero and one which allows for the focal points. Then they use the posterior density mean as the individual's estimated subjective survival probability. This mean will never be at the boundary of the interval from zero to one so that the adjusted subjective survival probabilities do not contain any more focal

[5] See Anderson (1999) and Armstrong (1998) for a discussion on how to construct complete annual U.S. life tables.

points.

2.4 Extensions and Results

Another problem with health curves as compared to mortality curves is that the 'state of death' is absorptive whereas having a work limiting health problem can be transitory. I therefore extend their original model to include subjective Markov switching probabilities, that are conditioned on gender, age and an individual's subjective health expectations

$$P_i\left(health\ state_{t+1} | health\ state_t,\ age_t,\ r.worklm_t\right) = \left(\begin{array}{cc} p_{i,hh} & p_{i,hs} \\ p_{i,sh} & p_{i,ss} \end{array} | age_{i,t},\ r.worklm_{i,t} \right),$$

where $p_{i,hh}$ is the conditional probability of having no work limiting health problem next year given that individual i has had no work limiting health problem this year, $p_{i,hs}$ is the subjective conditional probability of having a work limiting health problem next year given the individual has none this year, $p_{i,sh}$ is the probability of getting healthy next year given that the individual is sick this year and finally, $p_{i,ss}$ is the conditional transition probability of remaining work impaired due to lack of health.

In order to calculate the second row in the conditional Markov switching matrix we first construct reverse health tables that record the fraction of population per age group that recovers from a work limiting health problem. Since I do not have a corresponding question in our survey concerning expectations about recovering from health limitations I can only use realized probabilities of $p_{i,sh}$ and $p_{i,ss}$.

The paper is structured as follows. The next section describes the data. Section 4 describes the variable "work limiting health problems". Section 5 analyzes the subjective health expectations. I include tests for consistency and rationality of these health expectations. Section 6 develops the subjective health expectations curves, that can be used in structural estimations of life-cycle models. Section 7 brie□y discusses whether agents can learn their health expectations. Section 8 adds a discussion about why we should care about work limiting health problems. Finally, section 9 concludes the paper. Appendix A contains propositions from Gan, Hurd and McFadden (2003). Appendix B describes the algorithm that is used to compute the adjusted subjective health expectations curves. The rest of the appendices contain tables and figures.

3 The Data

I use seven waves of the RAND-HRS survey, 1992, 1994, 1996, 1998, 2000, 2002, and 2004. The RAND-HRS is developed from the health and retirement study (HRS) by the RAND Center of Aging. It is a composite data set that combines 4 cohort studies to get a national representative of the older population in the U.S. The cohorts are the AHEAD cohorts born before 1924, the CODA cohorts born between $1924 - 1930$, the HRS cohorts born between $1931 - 1941$ and the War Baby cohorts born between $1942 - 1947$. The largest of these surveys is the Health and Retirement Survey (HRS) conducted by the Institute for Social Research at the University of

Michigan. It is a longitudinal survey conducted every two years from 1992-2002. It covers a broad range of topics, including health, income, assets, employment, retirement, insurance, and family structure.

The majority of respondents in wave 1 of the HRS were 51 to 61 years old when the survey was first conducted in 1992. The baseline survey included $12,652$ persons, or $7,600$ households, with over samples of Mexican Americans, African Americans and residents of Florida. Juster and Suzman (1995) present a general overview of the HRS, Wallace and Herzog (1995) review the health measures in particular and Hurd and McGarry (1995) evaluate the subjective probabilities of survival. In the following I will concentrate on the population aged between 40 and 60 years in wave 1 and who will turn 52 and 72 years respectively in wave 7. Figure 1 contains histograms of the age distribution of all waves including a histogram of the age distribution over all waves. We see that the sample covers mostly individuals from age $45-75$. Table 7 reports the number of observations per wave including the number of reported deaths. Sample entries and exits other than deaths are not shown.

Wave 7 data do not contain the variable about expected work limiting health problems anymore. However, it still carries the variable measuring whether health limits the amount of work one can do. Summary statistics of expected work limiting health problems are therefore restricted to waves $1-6$.

4 Work Limiting Health Problems?

In this section I analyze the binary variable work limiting health problems of the RAND-HRS data set. I will denote this variable as *WorkLimHealthProblems* throughout the rest of the paper. The question wording in the HRS survey is:

"Now I want to ask how your health affects paid work activities. Do you have any impairment or health problem that limits the kind or amount of paid work you can do?"

In order to qualify this variable I run regressions of the form

$$WorkLimHealthProblems_{it} = \alpha_i + h_{it}\beta + x_{it}\gamma + \varepsilon_{it},$$

where h_{it} are health indicators and x_{it} are demographic variables (see Appendix C for regression results). Health indicator variables are self-reported health states (excellent, very good, good, fair, poor), the body mass index, indicators that measure the difficulties of daily activities like walking across the room, walking around the block, pushing large objects, sitting for more than 2 hours, using the phone, using money, climbing stairs, lifting 10 pounds, feeling depressed, having back problems. Furthermore I include doctor diagnosed health problems like high blood pressure, diabetes, cancer or tumors, lung problems, heart attacks and related heart problems, strokes, psychological problems, and arthritis and rheumatism. I also include measure of changes from last period in these diagnosed health problems. All indicators are binary variables except for the self-reported health state which is reported on a scale from 1 to 5, where 1 is excellent health and 5 is poor health.

Variable x_{it} is composed of demographic variables, lifestyle variables and income/expenditure variables. Demographic variables are age, gender, year of education, partnership status, whether parents are still alive.

Income/expenditure variables are total household income, individual earnings (of the head of the household), out-of-pocket medical expenses, total health expenditures, employment status, whether the job requires physical effort. Finally lifestyle variables contain whether the individual exercises and her smoking status.

I use standard OLS estimation, correcting for heteroskedastic errors. Since *WorkLimHealthProblems* is a binary variable, the linear probability model is not the best way to estimate this problem, although estimated coefficients are very easy to interpret. I therefore include nonlinear estimates from a Logit and a Probit model.

Tables 1 to 6 contain the results from panel regressions. We see that almost all coefficients in table 1 are positive and have $p-values$ smaller than 0.01 (indicated with three starts). That is standard health indicators for activities of daily living are highly correlated with work limiting health problems. Table 2 reports the correlation of work limiting health problems and doctor diagnosed health problems like high blood pressure, diabetes, cancer, lung diseases etc. In this case only diagnosed heart problems, psychological problems and arthritis have $p-values$ smaller than 0.05 for most model specifications (and positive signs).

Table 3 reports changes in doctor diagnosed health problems from the respective previous survey period and shows that most coefficients are negative as one would expect but again insignificant. Table 4 contains wealth measures that are mostly negatively correlated with health problems and significant, except for total household income. This makes intuitive sense, since wealth can be expected to be higher if an individual is healthy and can work. I would therefore expect a negative correlation of wealth and work limiting health problems.

From the demographic regressors in table 5 we see that women are less likely to develop health problems and that age is significant and positively correlated with work limiting health problems. Finally, lifestyle choices do have an effect, see table 6. Regular exercise is significantly negatively related with health problems. Smoking causes work limiting health problems but is insignificant in most regression specifications.

A word of caution is appropriate. The regressions in this section suffer from an endogeneity problem. There are unobserved factors that will influence both, work limiting health problems as well as the health indices that I use to describe them. In this case a regression measures only the magnitude of association and the direction of causation is not identified.

Note that some entries for the fixed effects Logit model are missing. This is due to lack of intragroup variation of that particular variable that we cannot estimate with a fixed effects estimator (e.g. gender, ever smoked, more than 12 years of education). In addition due to the construction of the fixed effects Logit model which drops observations without enough variation in the dependent variable when forming the conditional likelihood, the Logit model only uses 990 observations.[6]

I also test for fixed effects in the linear probability model using a Hausman test and cannot reject the hypothesis that estimates from the consistent (but possibly less efficient fixed effects estimator) are the same as the possibly inconsistent but more efficient random effects estimator. I therefore conclude that it is safe to use the more efficient random effects estimator.[7]

[6] See (Cameron and Trivedi, 2005, p. 800ff) for details.

[7] All non-linear panel estimates have incorrect standard errors and therefore the p-values are incorrect. Most computer packages report the wrong standard errors in panel estimation in the sense that they are based on restrictive distributional assumptions such as iid errors in the fixed effects models, and iid individual effects and iid errors in the random effects model. Stata 9 has already newer $xtreg$ commands that take care of this for the linear estimators. For the non-linear estimators we have to either panel bootstrap the standard errors or use a cluster-robust standard errors option if available. The problem we ran into when bootstrapping was a considerable time cost due to the size of our panel. In this version of the paper we therefore only report the standard p-values and point to the fact that standard errors might be

I also include estimates of the Hausman Taylor type that assumes that some variables are correlated with the individual fixed effect α_i but exogenous with respect to the error ε_{it}. I assume that all health indicator variables h_{it} are endogenous in this sense and then use the Hausman Taylor type estimator. Since the number of time varying exogenous covariates is larger than the number of time invariant endogenous covariates, identification is not a problem.[8] The values of this estimator are very similar to the random effects estimator.[9]

The standard criticism concerning the use of self reported data in this context is that individuals tend to answer that they do have work limiting health problems to justify that they are out of work. Estimates therefore tend to overstate the health effects on hours worked. See French (2003) for a discussion on this issue. Other issues with self-reported mortality and health data include perception differences by age and socio-economic status (e.g. Sen (2006), Crossley and Kennedy (2002)) as well as nationality (e.g. Jürges (2006)). Another issue concerns the context bias of survey answers. Burkhauser et al. (2002) use a health based survey and an employment based survey and compare the validity of self-reported work limiting health problems in tracking the prevalence of disability and employment of health impaired workers in a population. They find that although the measures fail to predict employment levels of health impaired workers they are able to capture the trends in employment. Differences in the trend outcomes achieved with survey responses from the health based versus the employment based survey tend to be insignificant.

5 Expectations about Work Limiting Health Problems

The variable concerning individual expectations about future work limiting health problems is denoted $r.worklm$ in the RAND-HRS. I will call this variable *ExpHealthProblems*. The exact wording of the survey question is:

"What about the chances that your health will limit your work activity during the next 10 years?".

Respondents can answer with a number from 0 to 100, where 0 indicates absolutely no chance of developing a work limiting health problem and 100 means that it is absolutely certain that a health problem will develop. Histograms of *ExpHealthProblems* for all waves are reported in figures 2 and 3. We clearly see that self-reported expectations show focal point responses, especially high at $0\%, 50\%$ and 100% chance of having a

underestimated for the non-linear models. See Cameron and Trivedi (2005), Chapter 21.

We experienced a similar computational time problem when calculating the marginal effects for the Logit and Probit models and therefore simply report the estimated coefficients. A rule of thumb that relates the coefficients from OLS estimates with the coefficients of Logit- and Probit models is:

$$\beta_{OLS} \simeq \frac{1}{4}\hat{\beta}_{Logit},$$

$$\beta_{OLS} \simeq \frac{1}{2.5}\hat{\beta}_{\Pr obit}.$$

The so adjusted parameter estimates can then be interpreted in the usual way.

[8] See (Cameron and Trivedi, 2005, p. 760-762) for more details on the IV estimator for the Hausman-Taylor Hybrid model.

[9] One can use this subset of data as proxy for observations that actually changed over the years of the survey. If we run the same regressions as above on this subset, we can exclude parts of the endogeneity problem since only the new variation (or the new health shocks and the according changes in the formation of health expectations) is taken into account and a given unobserved factor that led to the initial formation of health expectation is partly neutralized. Obviously, the change in work limiting health problems and the change in a health condition could be due to an unobserved factor that also changed. In this case the endogeneity problem remains. Running the regression on the subset of the sample that experienced a change in the work limiting health condition is therefore only a very crude way to weaken a potential endogeneity problem as it will only eliminate time invariant factors. The signs of most of the estimated coefficients remain unchanged. However, due to the smaller sample the statistical significance of most of the earlier results is lost.

work limiting health problem within the next ten years.

5.1 Are Health Expectations Consistent with Health Outcomes?

In the following I present summary statistics on individuals aged between $40-60$ years in wave 1 and compare their expectations about work limiting health problems to mortality expectations and health status across waves. Appendix B contains the tables. I divide the sample into subgroups by educational attainment and wealth and income quantiles.

Table 9 presents health expectations across all waves according to educational attainment. Comparing the mean expectation we see that in wave 1, college educated individuals have lower expectations about having work limiting health problems in the future than their less educated counterparts. College and above report a 34.23% probability versus 43.87%, 42.14% and 38.59% for less than high school, GED and high school graduates respectively. This pattern is repeated across all waves, although in later waves as the population gets older the expectations of higher educated individuals moves closer to expectations of lower educated groups.

Table 10 compares health expectations to mortality expectations of smokers and non-smokers. I again find the consistent pattern that smokers expect health problems with a higher probability than non-smokers and have lower expectations about living to age 75 (liv75) and 85 (liv85) respectively.

Table 11 and table 12 summarize health expectations according to wealth and income quantiles. I find a similar convergence pattern as in the classification by educational attainment. Individuals in high wealth and income quantiles have lower subjective probabilities of having a work limiting health problem within the next ten years. As the population gets older the expectations converge somewhat for both wealth and income quantiles.

In table 13 I compare expectations about work limiting health problems and mortality expectations from wave 1 and wave 2. I find that 52.12% of individuals who responded in both wave 1 and wave 2 had higher expectations about contracting health problems in the future in wave 1 than in wave 2. On the other hand 28.55% increased their subjective probability of having health problems in wave 2, whereas 19.33% did not revise their health expectations from wave 1 to wave 2.

The same comparison for subjective life expectancies reveals that 40.59% decrease their subjective belief of living to age 75 from wave 1 to wave 2, whereas 44.56% decreased their belief of living to age 85. Roughly 15% give focal point responses in both waves for health expectations, whereas focal point responses for $liv75$ and $liv85$ are around 23% and 13% respectively. It might be surprising to find that a large fraction of respondents, 52.117% find it more likely to contract health problems when they are younger. On the other hand one could argue that an older agent who is closer to retirement and does not have any work limiting health problems will find it more likely to also not have any problems during the next 10 years. In this sense the numbers in table 13 do make sense. The large fraction of people, 40.59% and 44.56%, whose survival expectations up to a target age go down as they get older might be explained by additional health related information that comes into play. On the other hand, one would expect somebody who is older, say 67 and closer to a target life expectancy of, say, 75, would think to have a higher probability of living to that age than somebody who is two years younger. Tables 14 and 15 compare wave 2 to wave 3 and wave 3 to wave 4 respectively. I do not have observations of

$liv85$ for wave 5 and wave 6.

In table 16 I report summary statistics according to health status. The first panel in the table reports the proportions of individuals having a specific health status in wave 1 and wave 2. We see that 54.6% of people with excellent health in wave 1, do still report excellent health for wave 2, whereas 33.4% report their health status as very good and 0.2% report a decline in their health to the status of poor. Similarly, of the people with very good health in wave 1, 54.4% still have very good health in wave 2. In addition, 16.3% of those with very good health in wave 1 improved their health to the status of excellent in wave 2, whereas 25.1% saw their health decline to status "good". We see that health states are very persistent in the sense that for almost all health states 50% of the individuals remain in that stage.

Panel two in table 16 summarizes the mean expectations about work limiting health problems by health status. I find that individuals with better health status in both waves have lower expectations about future health problems. Individuals who could improve their health over the waves report lower subjective probabilities of future health problems. See panel 3 and the negative numbers in the upper right corner, where changes in expectations about future health are negative. Panel 4 and Panel 5 report the mean expectations of living to age 75 and age 85 respectively. We again see that individuals with a better health status report higher probabilities of surviving up to a target age.

Table 17 compares wave 1 and wave 6 expectations according to health status. The variable $liv85$ is not available for wave 5 and wave 6. The persistence of health status over six waves is still quite strong. Although fewer individuals can maintain a health status of excellent over all six waves, 36.4%. Comparing panels 4 in table 16 and table 17 we can see that people with the same health status in wave 6 have higher expectations to live to target age 75. This is what one would expect, given that these individuals are much older now, some of them probably very close to target age 75.

5.2 Describing the Validity of Health Expectations

In order to test for the validity of *ExpHealthProblems* I run the following test that is similar to Hurd and McGarry (2002) and Bloom et al. (2006) with some notable exceptions concerning the interpretation.

I run six separate regressions of the following form

$$WorkLimHealthProblems_t = \beta \times ExpHealthProblems_1 + X_1\gamma + \varepsilon,\ t = 2, ..., 7.$$

I regress the realization of work limiting health problems from wave 2 to wave 7, $WorkLimHealthProblems_j$, $t = 2, ..., 7$ on $ExpHealthProblems_1$ in wave 1. $WorkLimHealthProblems_j$ is equal to 1 if the individual reports to have health limiting work problems or if the individual died.

The question regarding the subjective expectations about work limiting health problems that was initially asked in year 1992, was intended to measure the long-term probability of falling out of "good" health (or staying in "good" health) over the next 10 years. I would therefore expect a somewhat stable predictive performance of $ExpHealthProblems_1$ on realizations of $WorkLimHealthProblems_j$ for the following 10 years. The effect should also be positive, so that a higher expected probability of acquiring health problems should result in a higher realization of such problems within the next 10 years. This is exactly what we can observe in tables 21, 22, 23, 24, and 25.

Table 21 presents a Probit regression of the realization of work limiting health problems on the expectations variable in wave 1. We see that the $ExpHealthProblems_1$ has a stable relationship with the dependent variable $WorkLimHealthProblems_j$ and is highly significant in all waves. However, in the final wave 7, that is 12 years after the initial interview the predictive power of the expectations variable deteriorates. Table 22 provides results from Probit regressions with additional covariates from wave 1, that I summarize here as X_1. It contains factors that are likely to have an in uence on the occurrence of work limiting health problems such as gender, age, disease conditions, health status, smoking behavior, income and wealth. Introducing these additional covariates we see that the effect of $ExpHealthProblems_1$ becomes much smaller, but is still significant. We also still observe the stable relation over all waves except for the last wave, wave 7. For wave 7, the wave that lies beyond the original projection horizon (the question only asks for expectations about the next 10 years), we again observe a drop in predictive power of the expectations variable.

I next use a linear probability model estimated with an instrumental variable estimator to correct for endogeneity of $ExpHealthProblems_1$. After instrumenting $ExpHealthProblems_1$ using 12 dummy variables constructed from parental age when alive or parental age at the time of parent's death, the results still confirm what I have found so far.[10] A valid instrument for health expectations is a variable that helps predict the outcome of health problems only via expectations of health problems. The assumption that parental mortality does just that is a strong assumption, since genetic factors are likely to in uence or predict the occurrence of health problems directly. However, I condition on current health status, which should control for effects of family background on previous health conditions, so that family background is a reasonable instrument (see Fang et al. (2007) for a discussion on how expectations can in uence outcome variables which they call the "Mickey Mantle Effect").

Table 23 contains the results of the linear probability model including the list of covariates X_1.[11] Table 24 contains the IV-Probit estimates. We can still observe that the predictive power of $ExpHealthProblems_1$ is similar from wave 2 to wave 6 but then drops off at wave 7. The estimates are all significant and larger than in the non-instrumented Probit. Finally, table 25 contains results of the IV-Probit including covariates of wave 1. The results show a similar pattern, but are not significant anymore.

[10] Following Bloom et al. (2006) we create twelve dummy variables that we use as instruments for $ExpHealthProblems_1$: Parent alive and mother's age < 75, parent alive and father's age < 75, parent alive and mother's age $75 - 85$, parent alive and father's age $75 - 85$, parent alive and mother's age > 85, parent alive and father's age > 85, parent deceased and mother's age of death < 50, father's age of death < 50, mother's age of death $66 - 75$, father's age of death $66 - 75$, mother's age od death > 75 and father's age of death > 75.

[11] We report three test statistics for the linear IV-estimation. The Durbin-Wu-Hausman test does not reject the H_0 that the OLS estimates are inconsistent.

The null hypothesis of the Durbin-Wu-Hausman test states that an ordinary least squares (OLS) estimator of the same equation would yield consistent estimates. A rejection of the null indicates that the effects of the endogenous regressors are meaningful and that an instrumental variables technique is required.

The p-value from the Sargan overidentification test indicates that the instruments are uncorrelated with the error term and therefore valid instruments.

We actually report the Hansen J-test which is a generalization of the Sargan test. The Hansen test becomes the Sargan test under conditional homoskedasticity. The H_0 is that instruments are exogenous, or valid. If the test statistic for overidentifying restrictions is large then the IV estimator is inconsistent, so that rejection of H_0 is interpreted as evidence that the instruments are endogenous. In our case we cannot reject the H_0. See (Cameron and Trivedi, 2005, p. 277) for more information on tests for exogeneity of instruments.

Finally the Cragg-Donald test for weak instrument statistic, Cragg and Donald (1993), shows that we do have weak instruments.

Stock and Yogo (2002) report critical values for the Cragg-Donald statistic for the presence of weak instruments based on two-stage least squares bias. Critical values are 20.69, 11.05, 6.06 and 4.32 for the 5%, 10%, 20% and 30% bias respectively. If the Cragg-Donald statistics is less than the critical value then the instruments are weak.

5.3 Testing the Informational Content of Health Expectations

To examine whether these health expectations carry useful information, I compare the subjective health probabilities with the actual occurrence of work limiting health problems a decade later. I report mean values of health expectations in wave 1 and wave 2 (*ExpHealthProblems)* and compare them to the realizations of health limiting problems in wave 6 (*HealthProblems)*. Results are reported in table 18 and table 19.

In table 18 it seems that health expectations are fairly inconsistent when compared with realized health problems approximately 10 years later. To see this compare $mean\,(ExpHealthProblems)$ in wave 1 with $mean\,(HealthProblems)$ in wave 6. However, if one accounts for individuals who either left the survey or died from wave 1 to wave 6 (I unfortunately cannot distinguish between the two cases) then health expectations seem fairly consistent.[12] Compare $mean\,(ExpHealthProblems)$ in wave 1 with $mean\,(ExpHealth\Pr oblemsA)$ in wave 6.

The same holds true in table 19 where I only include individuals without work limiting health problems in wave 1 and wave 2 respectively. It also appears that males slightly underpredict future health problems, whereas females slightly overpredict health problems.

From these summary statistics I conclude that expectations about future work limiting health problems are formed reasonably, that is consistent with later realizations of such health problems. I now discuss whether health expectations are formed rationally in a more formal framework.

5.4 Are Health Expectations Formed Rationally?

I employ the framework developed in Bernheim (1990), Benitez-Silva et al. (2003) and Benitez-Silva and Dwyer (2004) to test whether expectations about work limiting health problems are formed rationally.[13] An individual is trying to predict a variable X and has access to certain information during period t. I denote this information set by Ω_t. In period $t+1$ the information set is augmented by newly available information $\omega_{t+1,}$ so that the new information set is $\Omega_{t+1} = (\Omega_t, \omega_{t+1})$. In my model I impose that individuals form expectations according to

$$X_t^e = E\left(X|\Omega_t\right),$$

where E is the expectations operator. This guarantees that errors in expectations will be uncorrelated with the set of variables known at time t. It then follows that

$$E\left(X_{t+1}^e|\Omega_t\right) = E\left[E\left(X|\Omega_t,\omega_{t+1}\right)|\Omega_t\right] = E\left[X|\Omega_t\right] = X_t^e.$$

Benitez-Silva and Dwyer (2004) point out that in order for the above relation to hold it is essential to assume that new information – that is its conditional distribution, not just its mean – is correctly forecast. The evolution of expectations is

$$X_{t+1}^e = X_t^e + \eta_{t+1}, \tag{1}$$

[12] Smith, Taylor and Sloan (2001) also report that attrition between waves is approximately 20% that is not due to death. Adjusting for this they find that the death rates in the HRS data corresponds fairly well to the decennial life table measures.

[13] Pesaran (1987) contains an early critisim of the Rational Expectations Hypothesis and limits its use to steady state analysis.

where the expectations error is $\eta_{t+1} = X^e_{t+1} - E\left[X^e_{t+1}|\Omega_t\right]$ and

$$E\left[\eta_{t+1}|\Omega_t\right] = 0. \tag{2}$$

From expression (1) and (2) I can derive a regression framework to test for the rational expectations hypothesis, that is

$$X^e_{t+1,i} = \alpha + \beta X^e_{t,i} + \gamma\Omega_{t,i} + \epsilon_{t,i}, \tag{3}$$

where i indexes the individual, α is a constant, and γ is a parameter vector that estimates the effect of information in period t on period $t+1$ expectations. The rational expectations (RE) hypothesis then implies that $\alpha = \gamma = 0$ and $\beta = 1$ (strong RE). Weak rationality, according to Bernheim (1990), assumes $\gamma = 0$ and tests for $\alpha = 0$ and $\beta = 1$. In both cases expectations follow a random walk.

Running simple OLS regressions on (3) might be misleading due to measurement errors in the dependent variable. I already mentioned that there are focal point responses in the subjective expectations variables. These lead to trimodal error distributions instead of normal error distributions. Also, noisy self-reports and omitted variables can make estimation more complex. Individuals may exaggerate or underestimate their expectations or have other motives to misrepresent them.

If I run a simple regression without control variables (weak rationality assumes $\gamma = 0$) of the form

$$X^e_{t+1,i} = \alpha + \beta X^e_{t,i} + \epsilon_{t,i}, \tag{4}$$

we can see that estimates for β are not close to one at all and estimates for intercept α are significantly different from zero (see table 26, second column). From this I would conclude that health expectations are not formed according to our theory, so that I would have to reject the weak rationality hypothesis. The same holds true for strong rationality as can be seen in the first column of table 26. I now follow Bernheim (1990) who claims that one should instrument the *ExpHealthProblem* with other subjective expectations variables. The use of these variables as instruments is based on the assumption that individuals' expectations are internally consistent, in the sense that all expectations are based on the same information. I therefore use the mortality expectations $liv75$ and $liv85$ as instruments for work limiting health expectations, *ExpHealthProblem*.

Column 3 and 4 in table 26 report the regression results for strong rationality and weak rationality. We now see that the coefficients on *ExpHealthProblem* are indistinguishable from 1 and the intercepts are not significant. In the $IV-$regression testing the strong rationality assumption most regressors that stand for information matrix Ω_t are insignificant as well. This leads us to conclude that we cannot reject the rational expectations hypothesis anymore and that the expectations variable *ExpHealthProblem* follows a random walk. Tests of this kind have low power though, so that we have to interpret the results with care.

I also ran these tests for different age groups (e.g. $40-50$, $55-60$, and $60-65$) to see whether agents become more rational as they get older. I find that, indeed, the rational expectations hypothesis can only be rejected for the older cohorts.

6 Subjective Health Expectations Curves

In this section I derive subjective health expectations curves, using the methodology developed in Gan and McFadden (2005) to correct for focal point responses. Table 20 lists the percentage of those respondents who gave continuous responses, focal responses, and no responses in the first two waves. The table also reports transition probabilities of the different response modes over the first two waves. We see that in wave 1 only 41.76% of respondents gave continuous responses with 12.24% providing focal point responses. A relatively large section of respondents gave no answer to the expectations health question, 46.13%.

The focal point responses cannot represent respondents' true probabilities, so that without correcting for focal responses of zero or one, it is impossible to derive health curves that change over time. In this section I try to recover the "true" subjective health expectations curve for each respondent. I call these the adjusted subjective health expectations (curves).

The reason why focal point responses cannot reflect true probabilities is quite intuitive. If a respondent thinks that there is absolutely no chance, a zero probability, of having a work limiting health problem within the next 10 years, the question arises why one could not just take this value and postulate that the respondent will use exactly this expectation in her decision process. Since I ultimately want to model this decision process, why not work with this probability?

Health expectations that cover a decade cannot be made with absolute certainty. I assume that individuals know this when they actually make their optimizing decision and simply misreported their subjective probabilities. It therefore makes sense to correct this reporting error.

6.1 Construction of Health Tables using Population Health Hazard Rates

I first derive health tables for the U.S. using observed outcome probabilities from the data. Manski (1993) has already suggested that outcome probabilities can be used as proxies for subjective mortality expectations. I then update these tables using the subjective health expectations. The resulting adjusted subjective health expectations do not contain focal point responses anymore but contain the additional information carried in the observed outcome probabilities (health tables).

In order to construct the health tables I first define the hazard rates for having a work limiting health problem as

$$\lambda_0(t) = \Pr(T = t_j | T \geq t_j) = \frac{d(t)}{l(t)}, \tag{5}$$

where $d(t)$ is the number of individuals developing a work limiting health problem at age t and $l(t)$ is the total number of individuals aged t without a health problem at the beginning of the period. The number of individuals developing a work limiting health problem from age t to $t+1$ is

$$d(t) = -[l(t+1) - l(t)].$$

A period in this context is the two year interval between waves in the Rand-HRS survey. The zero subscript in (5) denotes that the variable is derived from population realizations and not from a specific individual.

In addition I can derive the "survival probability". Survival in this context means remaining without a work limiting health problem from one period to the next. I denote this survival, or better, health maintenance probability from birth, as

$$S_0(t) = \Pr[T \geq t] = \prod\nolimits_{j|t_j \leq t} (1 - \lambda_j) = \frac{l(t)}{l(0)},$$

where $l(t)$ is again the number of individuals aged t without work limiting health problems and $l(0)$ is the starting cohort of newly bournes.

The health table "survival probability" from age a up to t without censoring is

$$S_{0a}(t) = \frac{S_0(a+t)}{S_0(a)} = \frac{\frac{l(a+t)}{l(0)}}{\frac{l(a)}{l(0)}} = \frac{l(a+t)}{l(a)}.$$

The health table hazard rate is the negative of the percentage change in the survival probability or more formally

$$\lambda_0(t) = -\Delta \ln S_0(t) = -\frac{1}{S_0(t)} \dot{S}_0(t) = -\frac{d \ln(S_0(t))}{dt} = -\%\Delta S_0(t).$$

We can also express this as

$$\lambda_0(t) = -\frac{S_0(t+1) - S_0(t)}{S_0(t)} = -\frac{\frac{l(t+1)}{l(0)} - \frac{l(t)}{l(0)}}{\frac{l(t)}{l(0)}} = -\frac{l(t+1) - l(t)}{l(t)} = \frac{d(t)}{l(t)}. \tag{6}$$

The cumulative health-problem hazard function (in continuous time) is[14]

$$\Lambda_0(t) = \int_0^t \lambda_0(t)\, d\tau = \int_0^t -\frac{d \ln(S_0(t))}{dt} d\tau = -\ln S_0(t). \tag{7}$$

In figure 4 I report the health-hazard rates for men and women. I limit the sample to people who are 40 years of age and older. By assumption individuals start being at risk of a work limiting health problem at age 40. I then construct the Kaplan-Meier survival rate with 99% confidence bounds. I assume individuals live in good health (without work limiting health problems) until failure. Failure is defined as the onset of a work limiting health problem, given that no such prior condition existed. An individual who enters the survey with a health problem is assumed to have failed at the age of survey entry. An individual who recovers from a health problem and develops another health problem while still in the survey at a later age is counted again as having failed for that particular age group. An individual leaving the survey is a censored spell and decreases the number of individuals at risk without counting towards the number of failures.[15]

E.g. a 70 year old male entering the survey without a health problem and reporting a health problem at age $74, 76, 78$ is counted as having failed at age 74. If the same individual does not report a health problem at age 80, but again reports a problem at age 82, then a second failure is counted for the 82 year old age group.

[14] Compare also Venables and Ripley (2002) for formal details on hazard functions.

[15] See (Cleves, Gould and Gutierrez, 2004, p. 59-62) for a discussion of how to model repeated failures by the same individual in Stata's survival package. Compare also (Cameron and Trivedi, 2005, p. 580 - 584) for a brief introduction to non-parametric survival analysis.

Similarly, if a 64 year old female enters the survey with a health problem, she is assumed to have failed at age 64.

I then count the number of people at risk at each age $l(t)$ where $t = 40, ..., 95$. Individuals at risk are all individuals in the survey that have not yet left the survey and do not have a health problem. In this sense, individuals who recover from a health problem but are still in the survey, will reenter the set of people at risk. I then count the number of people who fail at each age t, that is people who report a health limiting work problem at at t. The hazard rate for age t to $t+2$ is then defined as

$$\lambda(t) = \frac{d(t)}{l(t)} \equiv \lambda(t).$$

Since the hazard rates are very volatile I fit a 5^{th} order polynomial with least squares to smooth out the edges. From the top panel in figure 4 we see that the health hazard rates for men are higher than those for women over almost the entire age range. I will later report estimation results based on the original hazard rates and on the smoothed versions. I find that the results are robust and do not depend on whether I smooth the hazard functions before applying the Bayesian updating procedure. In figure 5 I also report unconditional hazard rates that I have calculated assuming that a person with a work limiting health problem in consecutive years is counted as having failed multiple times. The previous hazard rates would only count a transition from a healthy state to a sick state as failure which would then be reflected in the hazard rate. If we count both transitions from healthy to sick and from sick to sick as failure then the resulting hazard rate will increase as we can see in figure 5.

In figure 6 I report the reverse direction, that is the "hazard" rates of recovering from a work limiting health problem. We see that these rates go down as the individual ages.

6.2 Subjective Hazard Rates and Survival Functions

We next turn our attention to the individual. The personal health-survival probability from age a to target age $a+t$ for individual i is $S_{ia}(t)$. Variable $S_{ia}(t)$ is a random variable and s_{iat} is a realization of this variable.[16] The density of random variable $S_{ia}(t)$ is $\pi(s_{ia}(t))$ or $\pi(s_{iat})$. The personal health-problem hazard rate at age a is denoted $\lambda_{ia}(t)$ and the cumulative hazard rate is $\Lambda_{ia}(t)$.

From (7) I can derive an individual $i's$ health "survival" probability (or health curve) as

$$S_{ia}(t) = \exp\left(-\Lambda_{ia}(a+t) + \Lambda_{ia}(a)\right) = \exp\left(-\int_0^t \lambda_{ia}(a+r)\,dr\right). \tag{8}$$

I next use an individual's response to the health related question in the interview asking for a probability of having a work limiting health problem within the next ten years. I denote this probability as $1 - p_{ia\tau}$, where i denotes the individual, a is the individual's age and τ is time. Then the survival probability, that is the probability of maintaining the good health status is $p_{ia\tau}$ and its density is conditional on the personal survival

[16] We closely follow Gan, Hurd and McFadden (2003) and adopt their notation.

probability from age a to age $a+t$ as in

$$f\left(p_{ia\tau}|S_{ia\tau}=s_{ia\tau}\right).$$

The method employed uses the population hazard function $\lambda_{0a}\left(a+t\right)$ as a base and modifies it to calculate individual hazard rates $\lambda_{ia}\left(a+t\right)$ according to the following hazard scaling function

$$\lambda_{ia}\left(a+t\right)=\gamma_i\lambda_{0a}\left(a+t\right), \tag{9}$$

where $\gamma_i>1$ indicates a "pessimistic" and a $\gamma_i<1$ an "optimistic" individual.[17]

With focal responses and response errors present in $p_{ia\tau}$ the personal survival curve is not forced through $p_{ia\tau}$ at $a+\tau$. In this case I employ a Bayesian approach to update the individual survival curve. I denote the prior belief about the personal survival curve density as $\pi\left(s_{iat}\right)$. The mean of the prior density is $\exp\left(-\Psi\Delta\Lambda_{0at}\right)$ and its standard deviation is σ_2. Parameter Ψ measures the population's average optimistic degree. Given S_{iat}, the self-reported survival probability p_{iat} has density $f\left(p_{iat}|s_{iat}\right)$ so that the difference between the survival probability S_{iat} and the self-reported survival probability p_{iat} is the measurement error. I use the observed $p_{ia\tau}$ to update the prior density $\pi\left(s_{ia\tau}\right)$ in order to obtain the posterior density $\pi\left(s_{ia\tau}|p_{ia\tau}\right)$. The posterior density is given by

$$\pi\left(s_{ia\tau}|p_{ia\tau}\right)=\frac{f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)}{\int f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)ds_{ia\tau}},$$

with mean μ_{ia} and standard deviation σ_1. It can be shown that the best estimator for $S_{i\tau}$ with a quadratic loss function $L\left(S_{it},\hat{S}_{it}\right)=E\left[S_{it}-\hat{S}_{it}\right]^2$ is the conditional expectation, so that

$$\hat{S}_{i\tau}=E\left(S_{i\tau}|p_{ia\tau}\right).$$

I then apply $\hat{S}_{i\tau}$ to the observed record of realized health problems to obtain the model's parameters σ_1,σ_2 and Ψ. The log-likelihood function is given as

$$\ln L=\sum_{\text{NoHealthProblems}}\ln\hat{S}_{it}+\sum_{\text{HealthProblems}}\ln\left(1-\hat{S}_{it}\right). \tag{10}$$

I next make some assumption concerning the prior distribution of random variable S_{iat}. I denote the distribution of S_{iat} as $\pi\left(s_{ia\tau}\right)$ and define it as a truncated normal distribution. The mean of S_{iat} is $\exp\left(-\Psi\Delta\Lambda_{0at}\right)$, the variance is σ_2^2 and the truncation is between $0<s_{ia}<1$. The prior distribution is

$$\pi\left(s_{ia};\Psi\right)=\frac{\frac{1}{\sigma_2}\phi\left(\frac{s_{ia}-v_{ia}}{\sigma_2}\right)}{\Phi\left(\frac{1-v_{ia}}{\sigma_2}\right)-\Phi\left(-\frac{v_{ia}}{\sigma_2}\right)},$$

[17] Gan, Hurd and McFadden (2003) also calculate an age scaling model which leads to inferior results. We therefore concentrate on the hazard scaling version of their model.

where v_{ia} is the mean and σ_2 the standard deviation of the normal distribution. Both values satisfy

$$\exp\left(-\Psi \Delta \Lambda_{0a\tau}\right) = v_{iat} - \sigma_2 \eta\left(0, 1, v_{iat}, \sigma_2\right).$$

The right hand side is the mean of the truncated normal according to the formula in the appendix.

The conditional density of the responses to interview survival questions is assumed to follow a censored normal distribution

$$\begin{aligned}
f\left(p_{ia\tau}|s_{ia\tau}\right) &= \phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}}{\sigma_1}\right) \text{ when } 0 < p_{ia\tau} < 1, \\
\Pr\left(p_{ia\tau}=0|s_{ia\tau}\right) &= 1-\Phi\left(\frac{\mu_{ia\tau}}{\sigma_1}\right), \text{ and} \\
\Pr\left(p_{ia\tau}=1|s_{ia\tau}\right) &= 1-\Phi\left(\frac{1-\mu_{ia\tau}}{\sigma_1}\right),
\end{aligned}$$

with variance σ_1^2. The expected value $E\left[S_{ia}\right]$ of the conditional distribution is

$$s_{ia} = 0 \times \Pr\left(p_{ia\tau}=0|s_{ia\tau}\right) + E\left[x|0<x<1\right] \times f\left(p_{ia\tau}|s_{ia\tau}\right) + 1 \times \Pr\left(p_{ia\tau}=1|s_{ia\tau}\right),$$

so that

$$s_{ia} = \left[\Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right) + \Phi\left(\frac{\mu_{ia}}{\sigma_1}\right) - 1\right]\left[\mu_{ia} - \sigma_1 \eta\left(0,1,\mu_{ia},\sigma_1\right)\right] + \left[1-\Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right)\right],$$

where it can be shown (see Appendix A) that $E\left[x|0<x<1\right] = \left[\mu_{ia} - \sigma\eta\left(0,1,\mu_{ia},\sigma_1\right)\right]$. Finally, given $p_{ia\tau}$, the posterior distribution is given by

$$\pi\left(s_{ia}|p_{ia\tau}\right) = \frac{f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)}{\int f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right) ds_{ia\tau}}.$$

Then the best estimator for S_{ia} under a mean square loss function is its mean, that is

$$\hat{S}_{ia} = E\left[S_{ia}\right] = \int_0^1 s_{ia}\pi\left(s_{ia}|p_{ia\tau}\right) ds_{ia} = \frac{\int_0^1 s_{ia}\phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right) ds_{ia}}{\int \phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right) ds_{ia\tau}}.$$

I get similar results for the focal point responses at $p_{iat} = 0$ and 1 so that I summarize the predicted survival probabilities as

$$\hat{S}_{ia} = \begin{cases} \dfrac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}, \text{ if } p_{iat}=0 \\ \\ \dfrac{\int_0^1 s_{ia}\phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)ds_{ia}}{\int\phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)ds_{ia\tau}}, \text{ if } 0<p_{iat}<1 \\ \\ \dfrac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}, \text{ if } p_{iat}=1. \end{cases} \tag{11}$$

Since respondents are interviewed every two years I can update the predictions according to whether they are still without work limiting health problems. Then the likelihood function changes from (10) to

$$\ln L = \sum_{\text{NoHealthProblems}} \ln \hat{S}_{ia2} + \sum_{\text{HealthProblems}} \ln\left(1-\hat{S}_{ia2}\right). \tag{12}$$

From (8) and (9) one can calculate the optimism parameter γ_i as

$$\begin{aligned} \hat{S}_{ia}(t) &= \exp\left(-\int_0^t \hat{\gamma}_i\lambda_{0a}(a+r)\,dr\right), \\ &\rightarrow \hat{S}_{ia}(t) = \exp\left(-\hat{\gamma}_i\Delta\Lambda_{0a}(t)\right). \end{aligned}$$

Taking logs I can solve for $\hat{\gamma}_i$ as

$$\hat{\gamma}_i = -\frac{\ln\hat{S}_{ia\tau}}{\Delta\Lambda_{0a\tau}},$$

so that

$$\hat{S}_{ia2} = \hat{S}_{ia\tau}^{\left(\frac{\Delta\Lambda_{0a2}}{\Delta\Lambda_{0a\tau}}\right)}. \tag{13}$$

Substituting (13) into the log-likelihood function (12) we have

$$\ln L = \sum_{\text{NoHealth Pr oblems}} \ln\hat{S}_{ia\tau}^{\left(\frac{\Delta\Lambda_{0a2}}{\Delta\Lambda_{0a\tau}}\right)} + \sum_{\text{Health Pr oblems}} \ln\left(1-\hat{S}_{ia\tau}^{\left(\frac{\Delta\Lambda_{0a2}}{\Delta\Lambda_{0a\tau}}\right)}\right). \tag{14}$$

For details on these derivations I refer to Gan, Hurd and McFadden (2003). I report the algorithm that solves this maximum likelihood problem in Appendix D.

6.3 Estimation Results

I use a subset of the data to estimate the likelihood function in expression 14. I only use wave 1 and wave 2 in order to contain the computation burden. I only keep observations where respondents report no work limiting health problem in wave 1. This reduces the data to 7001 observations, 3489 of which are males and 3512 of which are females.

I report estimation results for two separate models in table 32. The first is a restricted model where I set $\Psi = 1$ and estimate σ_1 and σ_2. In this case the mean of the prior distribution is equal to the realizations in the health-tables. I report standard errors in parenthesis. Standard errors where obtained using a Bootstrap routine on 500 subsamples with 400 observations each. The first column uses Health Table data using a 5th degree polynomial to smooth the Kaplan-Meier estimate of the survival curve. The second column uses the original Kaplan-Meier estimator for the health table survival curve. Finally, in column three I report the estimation results for the unrestricted model where parameter Ψ is also estimated. I find that $\hat{\Psi} = 2.37$ which indicates that individuals are much more pessimistic about their health than the objective realization rates in the health tables.

Finally, I construct the health curves using the estimates of the restricted model. The top panel of figure 7 displays the health survival probabilities (the probability of remaining without work limiting health problems) for a 50 year old man. The blue line depicts the survival rates of an individual claiming a 100 percent change of staying in good health (or a 0 percent chance of developing a work limiting health problem), whereas the red line is an individual stating a 0 chance of staying in good health within the next 10 years. The green line is the subjective survival rate of an individual with average expectations about her health. The solid black line is the health-table survival rate. Figure 8 displays the analog results for 60 year old individuals.

In addition, I plot the confidence bounds of the health table estimates. We see that the confidence bounds of the adjusted subjective health curves of individuals reporting $p_{iat} = 0$ or 1 lie well beyond the confidence bounds of the health table estimates. Therefore, a model using the health table realizations as proxies for subjective expectations neglects statistically significant information from subjective expectations.

Figures 9 and 10 plot the survival curves for the unrestricted model. We see that in this model agent are more pessimistic, which is reflected in the estimate of $\hat{\Psi} = 2.37$ and the lower subjective survival curves. I report the histogram of self reported health expectations, together with the histogram of self reported health expectations after adjusting for focal point responses using the restricted model and the unrestricted model in figure 11. We see that the focal point responses at 0 and 1 have disappeared and that the unrestricted model exhibits the more pessimistic subjective health expectations.

7 Can Agents Learn Health Expectations

I next investigate whether agents can learn their health expectations as they get older. I therefore plot the population health hazards rates (cumulative over 10 years) against the mean subjective expectations hazard for the next 10 years in figure 12. We can observe that for the age range of 40 to 75 year old agents the difference, measured as squared deviations, between the population realization rate and the subjectively expected rate is indeed decreasing, thereafter we observe a widening gap between health realizations and health expectations (see bottom panel of figure 12). I interpret this as weak evidence for learning. As agents become older, the difference between their health expectations and their health realizations decreases.

Also we observe that agents younger than 70 seem to be more pessimistic about their health than agents older than 70. This pessimism was already noted earlier when we estimated a relatively large hazard scaling parameter Ψ. In addition, Ludwig and Zimper (2007) find a similar pattern of pessimism of the young and optimism of the old when comparing mortality expectations.

In order to adjust for any generational effects I divide the sample into three birth year cohorts: 1907-1929, 1930-1945, and 1946-1964. The squared deviations between the population health hazard rates and the expected health hazard are plotted in figure 13. The decline in the discrepancy between health expectations and health realizations as agents get older can still be observed.

8 Why Do We Care About Work Limiting Health Problems and Expectations Formed About Them?

Work limiting health problems have significant effects on people's income and wage rates. See table 28 and table 29 for simple Mincer type regressions of income and wage rates on education- and experience type variables including industry and region specific dummy variables (not reported). I find that work limiting health problems are significantly negatively related to $\log(income)$ in all specifications. Work limiting health problems are furthermore significantly negatively related to $\log(wage)$ in all specification but the fixed effects panel regression.

Second, I analyze whether there is additional information in the subjective expectations about work limiting health problems *ExpHealthProblems*, that is not in expectations about mortality $r.liv75$ $(r.liv85)$.

Tables 30 and 31 report Probit and Logit regressions of *WorkLimHealthProblemW6* (wave 6 work limiting health problems) on wave 1 expectations of health problems, *ExpHealthProblemsW1* and mortality $r1liv75$ and $r1liv85$. I find that including mortality expectations into our regression model, the health expectations variable stays significant. This indicates that there is additional information in subjective health expectations that is not covered by subjective mortality expectations. Approximating health expectations by mortality expectations might neglect important information. I therefore consider it an improvement to use this health expectations information in lieu of the widely used mortality expectations when modeling health impairments (e.g. Hurd (1989) uses mortality expectations as health proxies).

To test the extent of that information one would have to incorporate subjective health expectations into consumption-savings models and compare their predictions to models using objective realizations of health states. Only then can one safely quantify the additional effect that subjective expectations carry. Modelling a life-cycle model and calibrating or estimating it would go beyond the scope of this paper and is left for future research. Gan et al. (2004), however, do find significant improvements in using subjective survival expectations. This should give an indication that a similar result is possible using subjective health expectations.

9 Conclusion

In this paper I use the framework in Gan, Hurd and McFadden (2003) and apply it to work limiting health expectations and the respective realizations of work limiting health problems. I then derive adjusted subjective health expectations curves. From these curves we can "read off" an individual's adjusted subjective health expectation conditioning on the individual's subjective health expectation as answered in the survey, her age and her gender. This information can be used in calibrating life-cycle models with health uncertainty but also in structural estimation procedures of the same type of models. It has been shown by Gan et al. (2004) that adjusted subjective expectations can improve estimation results significantly in structural estimations. In addition, I qualify the variables describing work limiting health problems and expectations about work limiting health problems. I find that subjective health expectations do contain additional information that is not incorporated in subjective mortality expectations and that the rational expectations assumption cannot be rejected for subjective health expectations. In addition, younger cohorts seem to be pessimistic about their health compared with outcome probabilities from constructed health tables.

References

Anderson, Robert N. 1999. “Method for Constructing Complete Annual U.S. Life Tables.” National Center for Health Statistics. Vital Health Stat 2(129).

Armstrong, Robert J. 1998. “Methodology of the National And State Life Tables: 1989-91.” U.S. Department of Health and human Services, 1(2).

Benitez-Silva, Hugo and Debra S. Dwyer. 2004. “The Rationality of Retirement Expectations and the Role of New Information.” SUNY-Stony Brook.

Benitez-Silva, Hugo, Debra S. Dwyer, Wayne-Roy Gayle and Thomas J.Muench. 2003. “Expectations in Micro Data: Rationality Revisited.” Working Paper 2003-059 University of Michigan Retirement Research Center.

Bernheim, Douglas B. 1990. How Do the Elderly Form Expectations? An Analysis of Responses to New Information. In *Issues in the Economics of Aging*, ed. David Wise. University of Chicago Press pp. 259–285.

Bloom, David E., David Canning, Michael Moore and Younghwan Song. 2006. “The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States.” NBER Working Paper 12688.
URL: *http://www.nber.org/papers/w12688*

Burkhauser, Richard V., Mary C. Daly, Andrew J. Houtenville and NigarNargis. 2002. “Self-Reported Work Limitation Data: What They Can and Cannot Tell Us.” Federal Reserve Bank of San Francisco.

Cameron, Colin A. and Pravin K. Trivedi. 2005. *Microeconometrics, Methods and Applications*. New York: Cambridge University Press.

Cleves, Mario A., William W. Gould and Roberto G. Gutierrez. 2004. *An Introduction to Survival Analysis Using Stata*. Revised edition ed. Stata Press.

Cragg, John G. and Stephen G. Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory* 9(2):222–240.
URL: *http://ideas.repec.org/a/cup/etheor/v9y1993i2p222-40.html*

Crossley, Thomas F. and Steven Kennedy. 2002. "The Reliability of Self-Assessed Health Status." *Journal of Health Economics* 21:643–658.

Elder, Todd E. 2007. "Subjective Survival Probabilities in the Health and Retirement Study: Systematic Biases and Predictive Validity." Working Paper WP 2007-159, Michigan Retirement Research Center, University of Michigan.

Eriksson, Ingebor, Anna Lena Unden and Stig Elofsson. 2001. "Self-Rated Health. Comparisons Between Three Different Measures. Resultsfroma Population Study." *International Journal of Epidemiology* 30:326–333.

Fang, Hanming, Michael P. Keane, Ahmed Khwaja, Marin Salm and Dan Silverman. 2007. "Testing the Mechanisms of Structural Models: The Case of the Mickey Mantle Effect." *American Economic Review Papers and Proceedings* forthcoming.

French, Eric. 2003. "The Effects of Health, Wealth, and Wages on Labor Supply and RetirementBehavior." Federal Reserve Chicago, Working Papers Series WB 2000-02.

Gan, Li, Guan Gong, Michael Hurd and Daniel McFadden. 2004. "Subjective Mortality Risk and Bequests." NBER Working Paper 10789.

Gan, Li, Michael Hurd and Daniel McFadden. 2003. "Individual Subjective Survival Curves." NBER Working Paper 9480.

Gan, Li, Michael Hurd and Daniel McFadden. 2005. *Analyses in Economics of Aging*. The University of Chicago Press chapter Individual Subjective Survival Curves, pp. 377–411.

Hurd, Michael D. 1989. "Mortality Risk and Bequests." *Econometrica* 57(4):779–813.

Hurd, Michael D. and Kathleen McGarry. 1995. "Evaluation of the Subjective Probabilities of Survival in the Health andRetirementStudy." *The Journal of Human Resources* 30 (Supplement)(Supplement):S268–S292.

Hurd, Michael D. and Kathleen McGarry. 2002. "The Predicitvie Validity of Subjective Probabilities of Survival." *The Economic Journal* 112:966–985.

Jürges, Hendrik. 2006. "True Health vs. Response Styles: Exploring Cross-Country Differences inSelf-ReportedHealth." MEA-University of Mannheim.

Juster, F. Thomas and Richard Suzman. 1995. "An Overview of the Health and Retirement Study." *Journal of Human Resources* 30(Supplement):S7–S56.

Ludwig, Alexander and Alexander Zimper. 2007. "A Parsimonious Model of Subjective survival Expectancy." Mannheim Research Institute for the Economics of Aging (MEA).

Manski, Charles F. 1993. "Dynamic Choice in Social Settings: Learnings from the Experiences of Others." *Journal of Econometrics* 58(2):121–136.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72(5):1329–1376.

Perozek, Maria G. 2005. "Using Subjective Expectations to forecast Longevity: Do Survey RespondentsKnow-Something We Don't Know?" Federal Reserve Board, Washington D.C.

Perry, Michael. 2005. "Estimating Life Cycle Effects of Survival Probabilites in the Health and Retirement Study." University of Michigan Retirement Research Center, Working Paper 2005-103.

Pesaran, Hashem M. 1987. *The Limits to Rational Expectations*. Blackwell.

Sen, Amartya. 2006. "Health: perception versus observation." *British Medical Journal* (324):860–861.

Smith, Kerry V., Donald H. Taylor and Frank A. Sloan. 2001. "Longevity Expectations and Death: Can People Predict Their Own Demise?" *American Economic Review* 91(4):1126–1134.

Stock, James H. and Motohiro Yogo. 2002. "Testing for Weak Instruments in Linear IV Regression." NBER Technial Working paper 284.
URL: *http://ssrn.com/paper=346941*

Venables, William N. and B. Ripley. 2002. *Modern Applied Statistics with S*. 4 ed. New York: Springer Verlag.

Wallace, Robert B. and Regula A. Herzog. 1995. "Overview of the Health Measures in the Health and Retirement Study." *Journal of Human Resources* 30(Supplement):S84–S107.

10 Appendix

10.1 Appendix A: Propositions[18]

Proposition 1 *(Mean of the truncated normal) If* $x\tilde{}N\left[\mu,\sigma^2\right]$ *and* e *and* f *are constant, then*

$$\begin{aligned} E\left[x|e\leq x\leq f\right] &= \mu-\sigma\eta\left(e,f,\mu,\sigma\right), \text{ where} \\ \eta\left(e,f,\mu,\sigma\right) &= \frac{\phi\left(\frac{f-u}{\sigma}\right)-\phi\left(\frac{e-u}{\sigma}\right)}{\Phi\left(\frac{f-u}{\sigma}\right)-\Phi\left(\frac{e-u}{\sigma}\right)}. \end{aligned}$$

Proposition 2 *(Mean of the censored normal) If* $x^{*}\tilde{}N\left[\mu,\sigma^2\right]$ *and*

$$x=\begin{cases} e & \text{if } x^{*}\leq e \\ x^{*} & \text{if } e\leq x^{*}\leq f \\ f & \text{if } f\leq x^{*} \end{cases},$$

where e *and* f *are constant, then*

$$E\left[x\right]=\Phi\left(\frac{e-\mu}{\sigma}\right)e+\left[\Phi\left(\frac{f-\mu}{\sigma}\right)-\Phi\left(\frac{e-\mu}{\sigma}\right)\right]\left[\mu-\sigma\eta\left(e,f,\mu,\sigma\right)\right]+\left[1-\Phi\left(\frac{f-\mu}{\sigma}\right)\right]f.$$

Proposition 3 *When* $p_{iat}=0$*, then*

$$\hat{S}_{ia}=\frac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}.$$

Proposition 4 *When* $p_{iat}=1$*, then*

$$\hat{S}_{ia}=\frac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}.$$

10.2 Appendix B: Algorithm

I would like to thank Li Gan for making Matlab code available to us. I next describe our implementation of the algorithm. This implementation differs from Gan's code in the sense that I needed to construct the outcome probabilities (recorded in Health Tables) first. I also restrict my attention to the hazard scaling model.

1. Construct health tables using the population realizations of the hazard rate λ for each age group a of the form
$$\lambda_{0a}\left(a\right)=\frac{d\left(a\right)}{l\left(a\right)}.$$

[18] We brie□y state the following propositions without proofs. Proofs can be found in Gan, Hurd and McFadden (2003).

2. Use individual data on subjective expectations about work limiting problems within the next 10 years, denoted as $ExpHealthProblems = (1 - p_{ia})$, so that the probability of NOT having a work limiting health problem is p_{ia}. I interpret this also as the perceived survival rate (survival in 'good health') of individual i at age a.

3. Create dummy variable $d_{i,a,a+2} = 1$ if individual i was in good health in period 1 at age a and is still in good health in period 2 at age $a+2$ and $d_{i,a,a+2} = 0$ otherwise.

4. Calculate the cumulative hazard rate $\Lambda_{0a}(a+10)$ up to the target age $a+10$. The target age is $a+10$ because p_{ia} is defined as the subjective belief about surviving 10 years without work limiting health problems. I use
$$\Lambda_{0a}(a+10) = \sum\nolimits_{t=1}^{10} \lambda_{0a}(a+t).$$

5. Calculate the cumulative hazard rate $\Lambda_{0a}(a+2)$ up to the next wave at age $a+2$ which is
$$\Lambda_{0a}(a+2) = \sum\nolimits_{t=1}^{2} \lambda_{0a}(a+t).$$

6. Likelihood Routine:

   (a) Solve for μ_{ia} out of
$$s_{ia} = \left[\Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right) + \Phi\left(\frac{\mu_{ia}}{\sigma_1}\right) - 1\right][\mu_{ia} - \sigma_1 \eta(0,1,\mu_{ia},\sigma_1)] + \left[1 - \Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right)\right]. \quad (15)$$
   Where s_{ia} is a grid vector from $[0, ..., 1]$ and therefor μ_{ia} is also a vector.

   (b) Solve for v_{iat} out of
$$\exp(-\Psi\Lambda_{ia}(a+10)) = v_{iat} - \sigma_2 \eta(0,1,v_{iat},\sigma_2). \quad (16)$$

   (c) Solve for $\hat{S}_{iat}$ distinguishing $p_{iat} = 0, 1$, or interior from (11).

   (d) Build log-likelihood function from
$$\ln L(\sigma_1,\sigma_2,\Psi) = \sum_{i=1}^{N}\left[d_{i,a,a+2} \ln \hat{S}_{ia\tau}^{\left(\frac{\Lambda_{ia}(a+2)}{\Lambda_{ia}(a+10)}\right)} + (1 - d_{i,a,a+2}) \ln\left(1 - \ln \hat{S}_{ia\tau}^{\left(\frac{\Lambda_{ia}(a+2)}{\Lambda_{ia}(a+10)}\right)}\right)\right].$$

   (e)
$$\left(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\Psi}\right) = \arg \max_{\{\sigma_1,\sigma_2,\Psi\}} \ln L\left(\sigma_1,\sigma_2,\Psi | \hat{S}_{iat}\right).$$
   The restricted model fixes $\Psi = 1$ and only estimates σ_1 and σ_2.

7. Construction of subjective health curves:

   (a) Given $\left(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\Psi}\right)$ solve for μ_{ia} and v_{iat} from (15) and (16).

(b) Calculate estimates for survival $\hat{S}_{at}\left(p_{at}=0\right), \hat{S}_{at}\left(p_{at}=\bar{p}\right)$ and $\hat{S}\left(p_{at}=1\right)$ from (11), where $\bar{p}$ is the average subjective probability of surviving in good health of the sample.

(c) Calculate the cumulative hazard rates from the hazard rates starting at a certain base age a so that

$$\begin{aligned} \Lambda_{0a}\left(a\right) &= \lambda_{0a}\left(a\right), \\ \Lambda_{0a}\left(a+1\right) &= \lambda_{0a}\left(a\right)+\lambda_{0a}\left(a+1\right), \\ &\vdots \\ \Lambda_{0a}\left(a+T\right) &= \sum\nolimits_{t=0}^{T}\lambda_{0a}\left(a+t\right). \end{aligned}$$

Then define the following vector

$$\Lambda_{0aT}=\left[\Lambda_{0a}\left(a\right), \Lambda_{0a}\left(a+1\right), ..., \Lambda_{0a}\left(a+T\right)\right].$$

So that the vector of survival rates in good health from age a to age $a+T$ is

$$S_{0aT}=\exp\left(-\Lambda_{0aT}+\lambda_{0a}\left(a\right)\right).$$

The addition of the initial hazard rate normalizes the survival function S_{0aT} to be equal to 1 at age a. The zero subscripts denote the fact that these are mortality rates and survival rates of the population and not of a particular individual. I denote vector S_{0aT} to be the health table (population) survival rate of an individual with age a up to age $a+T$.

(d) I finally update the health table survival rate with the subjective survival probability from the data $p_{ia\tau}$ using the hazard scaling model described earlier $\lambda_{ia}\left(a+t\right)=\gamma_i\lambda_{0a}\left(a+t\right)$. Where the estimate of γ for a particular individual i, aged a who answers with $p_{ia\tau}$ for the health expectations questions is

$$\hat{\gamma}_i\left(p_{ia\tau}\right)=-\frac{\ln\hat{S}_{ia\tau}\left(p_{ia\tau}\right)}{\Lambda_{a0}\left(a+10\right)},$$

where $\hat{S}_{ia\tau}\left(p_{ia\tau}\right)$ was calculated in step (b) above.

(e) The vector of subjective survival rates in good health is then

$$S_{iaT}\left(p_{ia\tau}\right)=\exp\left(-\hat{\gamma}_i\left(p_{ia\tau}\right)\overbrace{\left[-\Lambda_{0aT}+\lambda_{0a}\left(a\right)\right]}^{S_{0aT}}\right),$$

where I plot these rates for $p_{ia\tau}=0,1$ and $\bar{p}$ in figure 7 for $a=50$ and in figure 8 for $a=60$.

10.3 Appendix C: Regression Tables

$$\text{OLS: } WorkLimHealthProblems_{it} = \alpha_i + h'_{it}\beta + x'_{it}\gamma + \epsilon_{it}$$

$$\text{Probit: } Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \Phi(\alpha_i + h'_{it}\beta + x'_{it}\gamma)$$

$$\text{Logit: } Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \frac{exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}{1 + exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}$$

	FE-OLS	RE-OLS	RE-Logit	FE-Logit	RE-Probit	IV-HTaylor
	(1)	(2)	(3)	(4)	(5)	(6)
Health status: Very Good	-.006	-.009	.187	.058	.071	-.003
	(.011)	(.004)**	(.126)	(.301)	(.058)	(.008)
Health status: Good	.014	.027	.785	.591	.365	.026
	(.017)	(.006)***	(.126)***	(.319)*	(.059)***	(.010)**
Health status: Fair	.047	.085	1.009	.838	.510	.066
	(.032)	(.012)***	(.144)***	(.375)**	(.071)***	(.015)***
Health status: Poor	.198	.243	1.595	1.973	.862	.237
	(.077)**	(.027)***	(.218)***	(.596)***	(.114)***	(.029)***
Diff. walking accross room	.145	.147	.777	.682	.403	.146
	(.115)	(.041)***	(.295)***	(.542)	(.158)**	(.037)***
Body mass index	.00004	-.0003	.002	-.017	.001	.002
	(.004)	(.0006)	(.007)	(.042)	(.004)	(.002)
Diff. walking 1 block	.128	.143	.778	.724	.448	.136
	(.033)***	(.014)***	(.090)***	(.201)***	(.050)***	(.012)***
Diff. pushing large objects	.064	.102	.738	.528	.423	.070
	(.029)**	(.012)***	(.090)***	(.194)***	(.049)***	(.011)***
Diff. sitting 2 hours	.025	.051	.476	.336	.264	.028
	(.022)	(.009)***	(.081)***	(.180)*	(.043)***	(.009)***
Diff. using the phone	.012	.053	.414	-.457	.235	.014
	(.122)	(.040)	(.347)	(.638)	(.171)	(.041)
Diff. using money	.020	.041	.384	.0007	.208	.025
	(.047)	(.021)**	(.170)**	(.378)	(.094)**	(.021)
Diff. climbing stairs	.049	.039	.409	.849	.221	.053
	(.018)***	(.007)***	(.077)***	(.184)***	(.040)***	(.008)***
Diff. lifting 10 pounds	.025	.086	.580	-.015	.332	.034
	(.031)	(.013)***	(.093)***	(.201)	(.051)***	(.012)***
Felt depressed	.049	.027	.218	.613	.107	.049
	(.020)**	(.009)***	(.088)**	(.221)***	(.047)**	(.010)***
Back problems	.008	.030	.445	.041	.243	.010
	(.014)	(.006)***	(.071)***	(.184)	(.037)***	(.008)
Constant	.309	.322	-.101		-.180	.379
	(.354)	(.044)***	(.535)		(.281)	(.090)***
Number of observations	15844	15844	15843	1593	15843	15844
R^2	.087					
F-statistic	1.963					33.321

Table 1: Non Linear Panel Wave (1-6): WorkLimHealthProblems on Health Indicators

OLS: $WorkLimHealthProblems_{it} = \alpha_i + h'_{it}\beta + x'_{it}\gamma + \epsilon_{it}$

Probit: $Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \Phi(\alpha_i + h'_{it}\beta + x'_{it}\gamma)$

Logit: $Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \dfrac{exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}{1 + exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}$

	FE-OLS	RE-OLS	RE-Logit	FE-Logit	RE-Probit	IV-HTaylor
	(1)	(2)	(3)	(4)	(5)	(6)
DocDiag. high blood pressure	-.003	-.0004	.031	.001	.024	.016
	(.033)	(.006)	(.074)	(.444)	(.039)	(.017)
DocDiag. diabetes	-.069	.005	.050	-1.077	.022	-.039
	(.054)	(.012)	(.114)	(.730)	(.060)	(.028)
DocDiag. cancer/tumor	.007	.009	.114	.621	.059	.014
	(.064)	(.012)	(.135)	(.720)	(.071)	(.031)
DocDiag. lung problems	-.006	.032	.194	.146	.110	.039
	(.086)	(.018)*	(.148)	(.760)	(.079)	(.038)
DocDiag. heart attack	.048	.056	.452	.706	.263	.086
	(.057)	(.012)***	(.102)***	(.472)	(.055)***	(.024)***
DocDiag. stroke	-.093	.045	.396	-.875	.224	-.066
	(.147)	(.027)*	(.231)*	(.977)	(.124)*	(.054)
DocDiag. psych. problem	.047	.042	.265	.636	.167	.098
	(.075)	(.014)***	(.126)**	(.583)	(.068)**	(.030)***
DocDiag. arthritis/rheumatism	.044	.032	.383	.982	.212	.068
	(.029)	(.006)***	(.073)***	(.415)**	(.038)***	(.015)***
Constant	.309	.322	-.101		-.180	.379
	(.354)	(.044)***	(.535)		(.281)	(.090)***
Number of observations	15844	15844	15843	1593	15843	15844
R^2	.087					
F-statistic	1.963					33.321

Table 2: Non Linear Panel Wave (1-6): WorkLimHealthProblems on Doctor Diagnosed Health Problems

$$\text{OLS: } WorkLimHealthProblems_{it} = \alpha_i + h'_{it}\beta + x'_{it}\gamma + \epsilon_{it}$$

$$\text{Probit: } Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \Phi(\alpha_i + h'_{it}\beta + x'_{it}\gamma)$$

$$\text{Logit: } Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \frac{exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}{1 + exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}$$

	FE-OLS	RE-OLS	RE-Logit	FE-Logit	RE-Probit	IV-HTaylor
	(1)	(2)	(3)	(4)	(5)	(6)
Change in high blood pressure	-.002	.007	.075	-.076	.036	-.008
	(.030)	(.012)	(.147)	(.412)	(.079)	(.017)
Change in diabetes	.023	-.014	-.166	.661	-.084	.002
	(.056)	(.024)	(.246)	(.633)	(.127)	(.028)
Change in cancer/tumor	-.016	-.006	.006	-1.004	.019	-.021
	(.061)	(.028)	(.298)	(.741)	(.156)	(.032)
Change in lung problems	-.028	-.030	-.200	-.959	-.100	-.058
	(.089)	(.038)	(.333)	(.662)	(.174)	(.036)
Change in heart attack	-.041	-.046	-.225	-.516	-.125	-.062
	(.061)	(.024)**	(.202)	(.405)	(.109)	(.024)***
Change in stroke	.184	.020	-.095	1.541	-.088	.157
	(.152)	(.055)	(.498)	(.953)	(.255)	(.057)***
Change in psych. problem	.016	.012	.116	.014	.049	-.007
	(.071)	(.028)	(.246)	(.506)	(.128)	(.027)
Change in arthritis/rheumatism	-.013	-.027	-.277	-.220	-.153	-.021
	(.026)	(.011)**	(.136)**	(.327)	(.071)**	(.013)
Constant	.309	.322	-.101		-.180	.379
	(.354)	(.044)***	(.535)		(.281)	(.090)***
Number of observations	15844	15844	15843	1593	15843	15844
R^2	.087					
F-statistic	1.963					33.321

Table 3: Non Linear Panel Wave (1-6): WorkLimHealthProblems on Change in Doctor Diagnosed Health Problems

$$\text{OLS: } WorkLimHealthProblems_{it} = \alpha_i + h'_{it}\beta + x'_{it}\gamma + \epsilon_{it}$$

$$\text{Probit: } Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \Phi(\alpha_i + h'_{it}\beta + x'_{it}\gamma)$$

$$\text{Logit: } Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \frac{exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}{1 + exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}$$

	FE-OLS	RE-OLS	RE-Logit	FE-Logit	RE-Probit	IV-HTaylor
	(1)	(2)	(3)	(4)	(5)	(6)
Total houshold income1000	2.46e-06	2.26e-06	1.00e-05	.0005	4.96e-06	4.68e-06
	(3.80e-06)	(2.18e-06)	(.00004)	(.0003)*	(.00002)	(2.88e-06)
Individual earnings (in 1000)	-.0001	-.0003	-.012	-.004	-.006	-.0002
	(.0001)	(.00006)***	(.002)***	(.004)	(.001)***	(.00007)***
Out of pocked health expend.	-1.32e-06	-6.10e-07	-6.34e-06	-1.00e-05	-2.40e-06	-6.60e-07
	(1.50e-06)	(6.96e-07)	(8.38e-06)	(.00002)	(4.15e-06)	(5.91e-07)
Total medical expenditure	3.89e-07	2.80e-07	1.82e-06	2.54e-06	1.13e-06	4.21e-07
	(4.65e-07)	(1.57e-07)*	(1.15e-06)	(2.69e-06)	(6.35e-07)*	(1.26e-07)***
Employed	-.052	-.096	-.958	-.793	-.508	-.090
	(.019)***	(.007)***	(.078)***	(.225)***	(.041)***	(.007)***
Job requires physical effort	-.049	-.021	-.212	-.724	-.107	-.037
	(.022)**	(.006)***	(.078)***	(.250)***	(.040)***	(.007)***
Constant	.309	.322	-.101		-.180	.379
	(.354)	(.044)***	(.535)		(.281)	(.090)***
Number of observations	15844	15844	15843	1593	15843	15844
R^2	.087					
F-statistic	1.963					33.321

Table 4: Non Linear Panel Wave (1-6): WorkLimHealthProblems on Wealth Measures

OLS: $WorkLimHealthProblems_{it} = \alpha_i + h'_{it}\beta + x'_{it}\gamma + \epsilon_{it}$

Probit: $Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \Phi(\alpha_i + h'_{it}\beta + x'_{it}\gamma)$

Logit: $Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \dfrac{exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}{1 + exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}$

	FE-OLS	RE-OLS	RE-Logit	FE-Logit	RE-Probit	IV-HTaylor
	(1)	(2)	(3)	(4)	(5)	(6)
age	-.001	-.004	-.043	-.023	-.022	-.005
	(.002)	(.0005)***	(.006)***	(.032)	(.003)***	(.0008)***
Male		-.032	-.451		-.240	-.046
		(.006)***	(.090)***		(.046)***	(.009)***
Education > 12 years		.022	.317		.169	-.121
		(.006)***	(.080)***		(.041)***	(.037)***
Living with partner	-.003	-.0007	.087	-.237	.057	-.006
	(.048)	(.014)	(.175)	(.702)	(.094)	(.015)
Mother alive	.0003	.003	.020	.222	.009	.003
	(.022)	(.005)	(.071)	(.306)	(.037)	(.006)
Father alive	-.011	-.008	-.066	-.455	-.037	-.006
	(.027)	(.006)	(.101)	(.417)	(.051)	(.008)
ExpHealthProblem	.0006	.001	.015	.012	.008	.001
	(.0002)***	(.00009)***	(.001)***	(.003)***	(.0006)***	(.00009)***
Constant	.309	.322	-.101		-.180	.379
	(.354)	(.044)***	(.535)		(.281)	(.090)***
Number of observations	15844	15844	15843	1593	15843	15844
R^2	.087					
F-statistic	1.963					33.321

Table 5: Non Linear Panel Wave (1-6): WorkLimHealthProblems on Demographics

OLS: $WorkLimHealthProblems_{it} = \alpha_i + h'_{it}\beta + x'_{it}\gamma + \epsilon_{it}$

Probit: $Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \Phi(\alpha_i + h'_{it}\beta + x'_{it}\gamma)$

Logit: $Prob(WorkLimHealthProblems_{it} = 1|x_{it}, \beta, \gamma, \alpha_i) = \dfrac{exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}{1 + exp(\alpha_i + h'_{it}\beta + x'_{it}\gamma)}$

	FE-OLS	RE-OLS	RE-Logit	FE-Logit	RE-Probit	IV-HTaylor
	(1)	(2)	(3)	(4)	(5)	(6)
Vigorous physical activity	-.010	-.012	-.193	-.294	-.099	-.010
	(.010)	(.004)***	(.069)***	(.155)*	(.035)***	(.005)**
Ever smoked	.275	.002	.042	14.018	.018	.004
	(.335)	(.005)	(.078)	(1411.447)	(.040)	(.007)
Smokes now	-.029	-.005	-.066	-.334	-.035	-.010
	(.032)	(.007)	(.089)	(.337)	(.046)	(.008)
Constant	.309	.322	-.101		-.180	.379
	(.354)	(.044)***	(.535)		(.281)	(.090)***
Number of observations	15844	15844	15843	1593	15843	15844
R^2	.087					
F-statistic	1.963					33.321

Table 6: Non Linear Panel Wave (1-6): WorkLimHealthProblems on Life Style

Wave	Year	Number of Obs.	%	Died	%
1	1992	12,652	9.31	229	1.8
2	1994	19,871	14.62	1,061	5.3
3	1996	19,052	14.02	1,224	6.4
4	1998	22,608	16.64	1,321	5.8
5	2000	20,900	15.38	1,411	6.8
6	2002	19,577	14.40	1,106	5.6
7	2004	21,245	15.63	–	–
Total	–	135,905	100.00	6,352	

Table 7: Observations by Wave and Number of Deceased

10.4 Appendix D: Summary Statistics

Table 8: Summary by Expected Work Limiting Health Problem: Age 40-60

Variable	**Mean**	**Std. Dev.**	**N**
Total Sample			
age	54.284	4.237	57279
Total houshold income	281507.783	932087.692	49752
Individual earnings	22999.799	40320.862	49752
Years of education	12.548	3.106	57060
Male	0.597	0.49	57279
Not 0, Not 50, Not 100			
age	54.121	4.3	43078
Total houshold income	291801.145	938761.376	35551
Individual earnings	21498.651	43849.958	35551
Years of education	12.527	3.202	42881
Male	0.604	0.489	43078
ExpWorkLimHealth=50			
age	55.022	3.848	8465
Total houshold income	254789.334	1101410.793	8465
Individual earnings	27195.982	27840.88	8465
Years of education	12.707	2.711	8451
Male	0.569	0.495	8465
ExpWorkLimHealth=0			
age	54.192	4.236	4624
Total houshold income	271935.397	542662.273	4624
Individual earnings	27746.469	32832.934	4624
Years of education	12.642	2.827	4616
Male	0.599	0.49	4624
ExpWorkLimHealth=100			
age	55.367	3.82	1112
Total houshold income	195622.048	470782.79	1112
Individual earnings	19310.957	23905.788	1112
Years of education	11.797	3.123	1112
Male	0.548	0.498	1112

Table 9: Health Expectations by Educational Attainment (Age Group: 40-60 in Wave 1)

	LtHighSchool	GED	HighSchoolGrad	someCollege	CollegeAbove
Wave 1					
ExpHealthProblems					
Number of observations	1295	332	2230	1445	1379
Mean	43.876448	42.138554	38.591928	37.439446	34.234953
StDev	29.63378	27.525414	27.920115	27.761923	24.904285
Wave 2					
ExpHealthProblems					
Number of observations	1324	341	2325	1548	1424
Mean	40.627644	40.237537	36.224516	34.576227	32.176966
StDev	31.070853	28.866516	26.985197	28.846651	25.413231
Wave 3					
ExpHealthProblems					
Number of observations	793	236	1617	1099	1083
Mean	42.910467	40.694915	38.808287	38.411283	37.012927
StDev	31.664805	28.570036	28.985364	28.656627	26.123796
Wave 4					
ExpHealthProblems					
Number of observations	666	188	1408	928	927
Mean	41.728228	42.898936	40.610795	38.993534	36.992449
StDev	29.282339	27.984438	27.67525	27.235953	26.330532
Wave 5					
ExpHealthProblems					
Number of observations	524	162	1111	789	820
Mean	44.141221	47.734568	44.492349	41.779468	41.570732
StDev	29.170611	28.241152	26.668936	27.374022	25.216496
Wave 6					
ExpHealthProblems					
Number of observations	417	132	925	657	687
Mean	45.254197	51.05303	45.671351	42.659056	42.034934
StDev	30.329087	29.10951	27.433826	27.636403	27.262228

Table 10: Smoker and Non-Smoker Health Expectations (Age Group: 40-60 in Wave 1)

	NumSmokers	Mean	StDev	NumOfNonsmokers	Mean	StDev
Wave1						
ExpHealthProblem	1642	40.749086	28.292658	4666	37.558937	27.533459
ExpLive to 75	1642	62.411693	30.296621	4666	68.671239	26.57337
ExpLive to 85	1642	40.310597	32.321715	4666	47.038148	30.684465
Wave2						
ExpHealthProblem	1494	36.92905	29.148447	4959	35.037911	27.058088
ExpLive to 75	1494	65.617805	26.106905	4959	69.51462	23.875841
ExpLive to 85	1494	41.838688	31.310986	4959	45.831216	29.373332
Wave3						
ExpHealthProblem	924	40.501082	28.986772	3455	38.356295	27.880392
ExpLive to 75	924	66.831169	26.224501	3455	71.804052	24.086389
ExpLive to 85	924	41.928571	32.143491	3455	49.053546	30.487931
Wave4						
ExpHealthProblem	714	41.239496	27.641363	2888	38.740651	27.088322
ExpLive to 75	714	65.12465	26.136762	2888	70.82964	24.022188
ExpLive to 85	714	39.752101	31.653611	2888	47.154778	29.569266
Wave5						
ExpHealthProblem	537	45.013035	28.446694	2381	42.102058	26.497837
ExpLive to 75	537	61.837989	28.732454	2381	70.662747	24.434359
Wave6						
ExpHealthProblem	368	44.758152	27.23845	1709	42.282036	27.143992
ExpLive to 75	368	63.179348	29.684727	1709	70.034523	25.442604

Table 11: Health Expectations per Wealth Quantiles (Age Group: 40-60 in Wave 1)

	1st Quantile	2nd Quantile	3rd Quantile	4th Quantile
Wave 1				
ExpHealthProblems				
Number of observations	1301	1788	1811	1781
Mean	43.866257	40.190157	37.007178	34.941044
StDev	29.342182	27.354288	27.243316	26.940621
Wave 2				
ExpHealthProblems				
Number of observations	1383	1799	1936	1849
Mean	40.375271	37.394108	34.737603	32.890752
StDev	31.407805	28.368283	26.54356	26.497712
Wave 3				
ExpHealthProblems				
Number of observations	898	1311	1343	1283
Mean	43.297327	39.470633	37.341772	37.480125
StDev	30.24603	29.447905	27.69326	27.897576
Wave 4				
ExpHealthProblems				
Number of observations	770	1140	1147	1067
Mean	40.698701	39.642105	40.061029	38.63074
StDev	30.113658	27.526824	26.592585	26.943108
Wave 5				
ExpHealthProblems				
Number of observations	618	948	1001	843
Mean	46.454693	42.691983	43.535465	41.217082
StDev	28.294791	26.9657	26.601146	26.331733
Wave 6				
ExpHealthProblems				
Number of observations	515	791	796	722
Mean	45.508738	44.60177	44.070352	43.405817
StDev	30.085375	28.468405	26.792941	27.370679

Table 12: Health Expectations per Income Quantiles (Age Group: 40-60 in Wave 1)

	1stQuantile	2ndQuantile	3rdQuantile	4thQuantile
Wave 1				
ExpHealthProblems				
Number of observations	353	1843	2320	2165
Mean	41.586402	41.220836	39.262931	35.307159
StDev	29.364075	28.887752	27.716295	26.313206
Wave 2				
ExpHealthProblems				
Number of observations	1330	1436	2067	2134
Mean	41.080451	37.771588	34.859216	32.917994
StDev	31.020011	29.023901	27.340296	25.876701
Wave 3				
ExpHealthProblems				
Number of observations	758	423	1799	1855
Mean	42.217678	41.193853	39.017232	37.32938
StDev	30.801171	29.67007	29.43682	26.927704
Wave 4				
ExpHealthProblems				
Number of observations	694		126	1697
Mean	40.992795		38.071429	40.721273
StDev	29.18397		28.057136	27.801002
Wave 5				
ExpHealthProblems				
Number of observations	621		1199	1590
Mean	42.832528		46.589658	40.909434
StDev	27.562029		27.391463	26.217142
Wave 6				
ExpHealthProblems				
Number of observations	595		748	1481
Mean	46.583193		46.533422	42.27684
StDev	28.072256		28.475833	27.644662
StDev				

Table 13: Percentage of Individuals according to Wave 1 and Wave 2 Expectations. Column 1 lists the fraction of the population for which expectations in Wave 1 are larger than expectations in Wave 2. Column 2 contains the population fraction for which expectations in Wave 1 are smaller than expectations in Wave 2. Column 3 contains the fraction of individuals who did not adjust their expectations between Wave1 and Wave 2. Column 1,2 and 3 add up to 100 %. Column 4,5 and 6 (row 1) contains the fraction of individuals who reported a 0%, 50% and 100% probability of acquiring work limiting health problems within the next 10 years for both, wave 1 and wave 2. Column 4,5 and 6 (row 2 and 3) report same for life expectancies to age 75 and age 85 respectively. (Age Group: 40-60 in Wave 1)

	r1>r2	r1<r2	r1=r2	Total	r1=r2=0	r1=r2=50	r1=r2=100
ExpHealthProblem	52.12	28.55	19.33	100	5.28	9.73	.63
ExpLive to 75	40.59	33.13	26.28	100	2.3	9.25	9.46
ExpLive to 85	44.56	36.41	19.02	100	4.77	4.8	3.36

Table 14: Percentage of Individuals according to Wave 1 and Wave 3 Expectations. (Age Group: 40-60 in Wave 1)

	r2>r3	r2<r3	r2=r3	Total	r2=r3=0	r2=r3=50	r2=r3=100
ExpHealthProblem	37.89	39	23.11	100	6.05	11.63	.76
ExpLive to 75	37.58	34.72	27.7	100	2.06	10.49	8.9
ExpLive to 85	42.35	38.64	19.02	100	3.11	6.23	3.43

Table 15: Percentage of Individuals according to Wave 1 and Wave 4 Expectations. (Age Group: 40-60 in Wave 1)

	r3>r4	r3<r4	r3=r4	Total	r3=r4=0	r3=r4=50	r3=r4=100
ExpHealthProblem	63.03	22.03	14.94	100	3.63	8.22	.45
ExpLive to 75	58.66	21.21	20.13	100	1.38	7.24	7.08
ExpLive to 85	63.88	22.42	13.71	100	2.03	4.44	2.56

Table 16: Health Status in Wave 1 (row) and Wave 2 (column): (1) health Transition probabilities, (2) Mean work limiting health expectations by health status in wave 1 and wave 2, (3) Mean change in work limiting health expectations by health status in wave 1 and wave 2, (4) Mean mortality expectations to age 75 and (5) Mean mortality expectations to age 85. The column entries depict the health status in wave 2. (Age Group: 40-60 in Wave 1)

	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor
(1) Transition probabilites					
1 Excellent	.546	.163	.05	.035	.014
2 Very Good	.334	.544	.262	.084	.083
3 Good	.102	.251	.544	.345	.097
4 Fair	.017	.039	.128	.471	.444
5 Poor	.002	.004	.016	.065	.361
Total	1	1	1	1	1
(2) Mean health expectations					
1 Excellent	.256	.359	.408	.3	.7
2 Very Good	.295	.364	.39	.514	.533
3 Good	.369	.378	.444	.54	.486
4 Fair	.415	.403	.482	.529	.759
5 Poor	.067	.586	.488	.564	.631
(3) Mean change in health exp					
1 Excellent	-.016	-.097	-.122	.037	-.7
2 Very Good	.009	-.049	-.04	-.03	-.317
3 Good	-.039	.011	-.051	-.101	.114
4 Fair	.011	.079	.007	-.031	-.106
5 Poor	.4	.236	.13	.161	.058
(4) Mean live75 expectations					
1 Excellent	79.314	70.427	71.053	39.524	70
2 Very Good	74.235	70.258	66.116	61.765	70
3 Good	71.324	67.202	61.115	52.895	36.923
4 Fair	61.667	64.773	58.017	49.431	36.471
5 Poor	90	76	54.595	46.078	41.311
(5) Mean live85 expectations					
1 Excellent	52.013	52.013	52.013	52.013	52.013
2 Very Good	50.441	50.441	50.441	50.441	50.441
3 Good	50.441	50.441	50.441	50.441	50.441
4 Fair	34.667	34.667	34.667	34.667	34.667
5 Poor	56.667	56.667	56.667	56.667	56.667

Table 17: Health Status in Wave 1 (row) and Wave 6 (column): (1) health Transition probabilities, (2) Mean work limiting health expectations by health status in wave 1 and wave 2, (3) Mean change in work limiting health expectations by health status in wave 1 and wave 2, and (4) Mean mortality expectations to age 75. The column entries depict the health status in wave 6. (Age Group: 40-60 in Wave 1)

	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor
(1) Transition probabilites					
1 Excellent	.364	.111	.047	.014	.211
2 Very Good	.423	.477	.253	.079	.316
3 Good	.177	.338	.504	.36	.316
4 Fair	.03	.068	.183	.511	.158
5 Poor	.005	.006	.013	.036	
Total	1	1	1	1	1
(2) Mean health expectations					
1 Excellent	.229	.324	.393	.85	.4
2 Very Good	.287	.36	.389	.645	.817
3 Good	.289	.375	.425	.538	.65
4 Fair	.458	.432	.446	.521	.533
5 Poor	.375	.24	.513	.48	
(3) Mean change in health exp					
1 Excellent	.079	-.003	-.002	-.35	-.087
2 Very Good	.102	.074	.049	-.032	-.225
3 Good	.156	.129	.079	-.051	-.267
4 Fair	.102	.066	.118	.048	.3
5 Poor	.35	.55	.25	.08	
(4) Mean live75 expectations					
1 Excellent	82.935	72.308	74.839	87.5	15
2 Very Good	75.014	71.061	66.975	59.5	41.111
3 Good	74.933	69.142	58.899	52.537	25.833
4 Fair	73.704	67.813	63.167	46.977	50
5 Poor	80	61.429	44	36.25	

Table 18: Mean of Self-Reported Health Expectations in Wave 1 and Wave 2 and the Realizations of Health Problems 10 Years Later in Wave 6. HealthProblemsA counts all individuals having left the survey (due to death or attrition) as having a health problem, so HealthProblems=1 for such individuals.

	45-49	50-54	55-59	60-64	65-69	70-74	70-74
MALE							
Wave1							
Mean(ExpHealthProblem)	.4	.278	.366	.384	.404	.428	.518
Number of observations	9	23	91	1565	1477	726	111
Wave2							
Mean(ExpHealthProblem)	.138	.333	.337	.349	.377	.391	.474
Number of observations	4	15	49	840	1384	919	236
Wave6							
Mean(HealthProblems)	.2	.176	.194	.201	.213	.252	.333
Number of observations	5	17	72	1249	1080	481	69
Wave6a							
Mean(HealthProblemsA)	.556	.391	.385	.388	.415	.468	.541
Number of observations	9	23	91	1564	1473	726	111
FEMALE							
Wave1							
Mean(ExpHealthProblem)	.279	.31	.343	.386	.394	.438	.45
Number of observations	56	187	585	1491	1294	385	8
Wave2							
Mean(ExpHealthProblem)	.258	.315	.293	.337	.368	.4	.511
Number of observations	57	123	420	1355	1492	828	24
Wave6							
Mean(HealthProblems)	.096	.119	.196	.223	.227	.254	.2
Number of observations	52	160	495	1189	1016	256	5
Wave6a							
Mean(HealthProblemsA)	.214	.243	.339	.395	.378	.448	.5
Number of observations	56	185	584	1490	1292	384	8

Table 19: Mean of Self-Reported Health Expectations of Individuals without Work Limiting Health Problems in Wave 1 and Wave 2 and the Realizations of Health Problems 10 Years Later in Wave 6. HealthProblemsA counts all individuals that have formed a health expectation in wave 1 and that have left the survey (due to death or attrition) as having a health problem in wave 6, so HealthProblems=1 for such individuals. HealthProblemsB does the same for individuals having formed expectations in wave 2.

	45-49	50-54	55-59	60-64	65-69	70-74	70-74
MALE							
Wave1							
Mean(ExpHealthProblem)	.325	.295	.367	.363	.384	.415	.511
Number of observations	8	21	82	1428	1322	651	102
Wave6							
Mean(HealthProblems)	.2	.125	.182	.166	.18	.218	.308
Number of observations	5	16	66	1142	976	436	65
Wave6a							
Mean(HealthProblemsA)	.5	.333	.378	.36	.385	.436	.52
Number of observations	8	21	82	1427	1319	651	102
Wave2							
Mean(ExpHealthProblem)	.017	.271	.291	.308	.342	.359	.442
Number of observations	3	12	43	734	1203	794	185
Wave6							
Mean(HealthProblems)	.2	.154	.159	.143	.164	.214	.321
Number of observations	5	13	63	1053	890	398	56
Wave6b							
Mean(HealthProblemsB)	.429	.267	.311	.307	.339	.393	.494
Number of observations	7	15	74	1249	1135	560	83
FEMALE							
Wave1							
Mean(ExpHealthProblem)	.257	.302	.326	.373	.379	.415	.433
Number of observations	53	175	528	1388	1186	341	6
Wave6							
Mean(HealthProblems)	.082	.099	.17	.191	.205	.199	0
Number of observations	49	151	447	1107	933	226	4
Wave6a							
Mean(HealthProblemsA)	.208	.225	.317	.369	.359	.409	.333
Number of observations	53	173	527	1387	1184	340	6
Wave2							
Mean(ExpHealthProblem)	.223	.271	.273	.305	.326	.363	.398
Number of observations	52	108	378	1177	1274	700	17
Wave6							
Mean(HealthProblems)	.07	.083	.162	.166	.179	.19	0
Number of observations	43	145	437	1001	853	205	4
Wave6b							
Mean(HealthProblemsB)	.178	.179	.277	.319	.305	.365	.333
Number of observations	45	162	487	1201	1027	293	6

Table 20: Focal Responses and Continuous Responses about Work Limiting Health Expectations and the Transitions from Wave 1 to Wave 2

	continuous	0	100	NA	Total
Wave1					
Continuous	4078	655	141	479	5353
Continuous in percent	76.18	12.24	2.63	8.95	41.76
0	710	401	33	120	1264
0 in percent	56.17	31.72	2.61	9.49	9.86
100	167	35	48	38	288
100 in percent	57.99	12.15	16.67	13.19	2.25
NA	40	15	2	5856	5913
NA in percent	.68	.25	.03	99.04	46.13
Total	4995	1106	224	6493	12818
Total in percent	38.97	8.63	1.75	50.66	100

$$\boxed{\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta ExpectedWorkLimHealthProblems_{92})}$$

COEFFICIENT	(1) Wave2_mfx HealthProb94	(2) Wave3_mfx HealthProb96	(3) Wave4_mfx HealthProb98	(4) Wave5_mfx HealthProb00	(5) Wave6_mfx HealthProb02	(6) Wave7_mfx HealthProb04
ExpHealthProblem	0.00192***	0.00190***	0.00198***	0.00201***	0.00218***	0.000500***
	(0.00013)	(0.00015)	(0.00015)	(0.00016)	(0.00016)	(0.00012)
Observations	7238	7238	7238	7238	7238	7238

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Validity of Work Limiting Health Expectations from a Probit. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The regressor is expected work limiting health problems of wave1. Age: 40-60 in Wave 1.

10.5 Appendix E: Rational Expectations Tests and Income Regressions

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta ExpectedWorkLimHealthProblems_{92} + \gamma X_{92})$$

COEFFICIENT	(1) Wave2_mfx HealthProb94	(2) Wave3_mfx HealthProb96	(3) Wave4_mfx HealthProb98	(4) Wave5_mfx HealthProb00	(5) Wave6_mfx HealthProb02	(6) Wave7_mfx HealthProb04
ExpHealthProblem	0.000618***	0.000519***	0.000672***	0.000778***	0.000986***	0.000273**
	(0.00013)	(0.00015)	(0.00015)	(0.00016)	(0.00017)	(0.00013)
age	0.00138	0.00156	0.00363***	0.000911	0.00236**	-0.0000130
	(0.00094)	(0.0010)	(0.0011)	(0.0011)	(0.0012)	(0.00088)
female	-0.0381***	-0.0442***	-0.0255***	-0.0197**	-0.0167	-0.0291***
	(0.0080)	(0.0093)	(0.0094)	(0.010)	(0.010)	(0.0076)
black	-0.0103	-0.0243**	-0.0218**	-0.0156	-0.0123	-0.0214**
	(0.0090)	(0.010)	(0.011)	(0.012)	(0.012)	(0.0085)
partner	-0.0226**	-0.0211**	0.00423	-0.00939	-0.0178	-0.0229**
	(0.0096)	(0.011)	(0.010)	(0.012)	(0.012)	(0.0095)
Education > 12 years	0.000259	0.00436	-0.0118	-0.0101	-0.00653	-0.00978
	(0.0077)	(0.0088)	(0.0091)	(0.0097)	(0.010)	(0.0073)
Health status: Good	0.0944***	0.0882***	0.0936***	0.115***	0.109***	0.0379***
	(0.014)	(0.015)	(0.015)	(0.016)	(0.016)	(0.011)
Health status: Fair	0.214***	0.195***	0.189***	0.189***	0.156***	0.0281*
	(0.027)	(0.026)	(0.026)	(0.026)	(0.025)	(0.016)
DocDiag. stroke	0.0797*	0.0527	0.0226	0.0848*	0.0181	-0.0298
	(0.042)	(0.042)	(0.039)	(0.049)	(0.043)	(0.025)
DocDiag. arthritis/rheumatism	0.0297***	0.0563***	0.0594***	0.0534***	0.0739***	0.0121
	(0.0083)	(0.0097)	(0.010)	(0.011)	(0.011)	(0.0081)
Smokes now	0.0220**	0.0459***	0.0406***	0.0366***	0.0461***	0.0264***
	(0.0094)	(0.011)	(0.011)	(0.012)	(0.012)	(0.0095)
Individual earnings (in 1000)	-0.000884***	-0.000546**	-0.000589***	-0.000577**	-0.000684***	-0.000207
	(0.00022)	(0.00025)	(0.00022)	(0.00023)	(0.00023)	(0.00015)
Employed	-0.0883***	-0.0152	-0.0110	-0.00573	0.00577	-0.0375*
	(0.024)	(0.022)	(0.022)	(0.023)	(0.023)	(0.021)
Observations	7158	7158	7158	7158	7158	7158

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 22: Validity of Work Limiting Health Expectations from a Probit. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. All regressors are from Wave 1. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (these are not reported here). Age: 40-60 in Wave 1.

$$\boxed{\text{WorkLimHealthProblems}_j = \beta \times ExpectedWorkLimHealthProblems_{92} + X_{92}\gamma + Z_{92}\theta}$$

COEFFICIENT	(1) HealthProb94	(2) HealthProb96	(3) HealthProb98	(4) HealthProb00	(5) HealthProb02	(6) HealthProb04
ExpHealthProblem	0.00225	0.00766	0.00666	0.00411	0.00447	0.00462
	(0.0038)	(0.0049)	(0.0048)	(0.0046)	(0.0047)	(0.0037)
age	0.000575	-0.00163	0.000812	-0.000583	0.000553	-0.00194
	(0.0017)	(0.0022)	(0.0022)	(0.0021)	(0.0021)	(0.0017)
female	-0.0327***	-0.0305**	-0.0121	-0.0103	-0.00516	-0.0221**
	(0.0099)	(0.012)	(0.012)	(0.012)	(0.012)	(0.0097)
black	-0.0130	-0.0213	-0.0191	-0.0159	-0.0114	-0.0177
	(0.012)	(0.015)	(0.015)	(0.015)	(0.015)	(0.011)
partner	-0.0216**	-0.0153	0.0107	-0.00732	-0.0153	-0.0177
	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)	(0.011)
Education > 12 years	-0.00200	0.00126	-0.0135	-0.0104	-0.00802	-0.0114
	(0.0079)	(0.0099)	(0.0098)	(0.0096)	(0.0099)	(0.0081)
Health status: Good	0.0451	-0.0168	0.00825	0.0565	0.0526	-0.0111
	(0.041)	(0.052)	(0.052)	(0.049)	(0.050)	(0.040)
Health status: Fair	0.157**	0.0418	0.0694	0.110	0.0773	-0.0513
	(0.070)	(0.089)	(0.088)	(0.083)	(0.086)	(0.068)
DocDiag. stroke	0.109**	0.0494	0.0135	0.0844	0.00837	-0.0519
	(0.052)	(0.056)	(0.055)	(0.055)	(0.054)	(0.036)
DocDiag. arthritis/rheumatism	0.0306**	0.0442***	0.0501***	0.0488***	0.0701***	0.00223
	(0.013)	(0.017)	(0.016)	(0.016)	(0.016)	(0.013)
Smokes now	0.0224*	0.0310*	0.0271	0.0294*	0.0377**	0.0167
	(0.013)	(0.017)	(0.017)	(0.016)	(0.016)	(0.013)
Individual earnings (in 1000)	-0.000397***	-0.000202	-0.000201	-0.000273*	-0.000300**	-0.0000610
	(0.00014)	(0.00017)	(0.00015)	(0.00016)	(0.00015)	(0.00013)
Employed	-0.110***	-0.0358	-0.0315	-0.0153	-0.00770	-0.0496**
	(0.025)	(0.028)	(0.027)	(0.027)	(0.027)	(0.023)
Constant	0.0200	-0.0610	-0.217***	-0.135*	-0.219***	0.0779
	(0.068)	(0.080)	(0.083)	(0.079)	(0.080)	(0.065)
Observations	7158	7158	7158	7158	7158	7158
R^2	0.14	-0.13	-0.05	0.06	0.05	-0.14
P-val Durbin-Wu-Hausman	0.717	0.118	0.206	0.504	0.496	0.273
P-val Hansen J-Stat	0.896	0.639	0.265	0.328	0.322	0.558
Cragg-Donald statistic	0.956	0.956	0.956	0.956	0.956	0.956

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 23: Validity of Work Limiting Health Expectations from Linear Probability IV-Regression. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The instrumented regressor is expected work limiting health problems of wave1. Instruments are 12 indicator variables constructed from age of parents when alive or at death. Sample Age: 40-60 in Wave 1.

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta \times ExpectedWorkLimHealthProblems_{92} + Z_{92}\theta)$$

COEFFICIENT	(1) Wave2_mfx	(2) Wave3_mfx	(3) Wave4_mfx	(4) Wave5_mfx	(5) Wave6_mfx	(6) Wave7_mfx
ExpHealthProblem	0.0312***	0.0306***	0.0413***	0.0382***	0.0390***	0.0172*
	(0.0096)	(0.0092)	(0.0099)	(0.0095)	(0.0094)	(0.0099)
Observations	7238	7238	7238	7238	7238	7238

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24: Validity of Work Limiting Health Expectations from IV-Probit Estimation. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The instrumented regressor is expected work limiting health problems of wave1. Instruments are 12 indicator variables age of parents when alive or at death. Sample Age: 40-60 in Wave 1.

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta \times ExpectedWorkLimHealthProblems_{92} + X_92\gamma + Z_{92}\theta)$$

COEFFICIENT	(1) Wave2 HealthProb94	(2) Wave3 HealthProb96	(3) Wave4 HealthProb98	(4) Wave5 HealthProb00	(5) Wave6 HealthProb02	(6) Wave7 HealthProb04
ExpHealthProblem	0.00389	0.00427	0.00305	0.00260	0.00560	0.00139
	(0.0034)	(0.0037)	(0.0037)	(0.0039)	(0.0042)	(0.0031)
age	-0.0000609	-0.000310	0.00221	0.00000213	0.000116	-0.000684
	(0.0016)	(0.0017)	(0.0018)	(0.0019)	(0.0020)	(0.0015)
female	-0.0304***	-0.0353***	-0.0172	-0.0124	-0.00357	-0.0267***
	(0.0098)	(0.011)	(0.011)	(0.011)	(0.012)	(0.0088)
black	-0.0109	-0.0257**	-0.0237*	-0.0178	-0.00997	-0.0218**
	(0.012)	(0.013)	(0.013)	(0.014)	(0.015)	(0.010)
partner	-0.0199*	-0.0186	0.00717	-0.00882	-0.0142	-0.0209**
	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.010)
Education > 12 years	-0.00266	0.00263	-0.0121	-0.00984	-0.00847	-0.0101
	(0.0082)	(0.0090)	(0.0090)	(0.0093)	(0.010)	(0.0075)
Health status: Good	0.0280	0.0184	0.0457	0.0721*	0.0409	0.0224
	(0.037)	(0.040)	(0.041)	(0.042)	(0.046)	(0.033)
Health status: Fair	0.128**	0.102	0.133*	0.136*	0.0574	0.00588
	(0.063)	(0.070)	(0.069)	(0.073)	(0.077)	(0.057)
DocDiag. stroke	0.104**	0.0594	0.0241	0.0888*	0.00507	-0.0424
	(0.053)	(0.052)	(0.051)	(0.054)	(0.055)	(0.033)
DocDiag. arthritis/rheumatism	0.0266**	0.0526***	0.0589***	0.0525***	0.0673***	0.0101
	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)	(0.012)
Smokes now	0.0187	0.0386***	0.0352**	0.0328**	0.0352**	0.0240**
	(0.013)	(0.014)	(0.014)	(0.015)	(0.016)	(0.012)
Individual earnings (in 1000)	-0.000362***	-0.000275*	-0.000278**	-0.000305**	-0.000276*	-0.000130
	(0.00013)	(0.00015)	(0.00014)	(0.00015)	(0.00015)	(0.00011)
Employed	-0.114***	-0.0279	-0.0231	-0.0117	-0.0103	-0.0421*
	(0.025)	(0.025)	(0.024)	(0.026)	(0.027)	(0.022)
Constant	0.0197	-0.0603	-0.216***	-0.135*	-0.219***	0.0784
	(0.070)	(0.072)	(0.077)	(0.077)	(0.082)	(0.061)
Observations	7158	7158	7158	7158	7158	7158
R^2	0.09	0.07	0.10	0.09	0.01	0.01

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25: Validity of Work Limiting Health Expectations from IV-Probit-Twostep Estimation. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The instrumented regressor is expected work limiting health problems of wave1. Instruments are 12 indicator variables age of parents when alive or at death. Sample Age: 40-60 in Wave 1. All regressors are from Wave 1. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (not all are reported here). Sample Age: 40-60 in Wave 1.

$$\text{ExpHealthProblems}_{1996} = \alpha_i + \beta \times ExpHealthProblems_{i,1994} + \Omega_{i,1994}\gamma + \epsilon_{i,1994}$$

	strongRE	weakRE	IV-strongRE	IV-weakRE
	(1)	(2)	(3)	(4)
ExpHealthProblem	.241***	.339***	1.088***	1.077***
rExpLive to 75	-.068***			
rExpLive to 85	-.068***			
Smokes now	1.334			
Vigorous physical activity	-.768			
Mother alive	1.926**			
Father alive	-2.392**			
Health status: Very Good	1.764*		-1.927	
Health status: Good	4.539***		-2.022	
Health status: Fair	7.835***		-3.891	
Health status: Poor	13.170**		-8.090	
Diff. walking accross room	-3.566		-12.459	
Body mass index	-.024		-.104	
Diff. walking 1 block	-.278		-6.725**	
Diff. pushing large objects	.685		-.912	
Diff. sitting 2 hours	1.027		-1.008	
Diff. using the phone	8.404		5.034	
Diff. using money	4.353*		5.901**	
Diff. climbing stairs	1.177		-.591	
Diff. lifting 10 pounds	5.006***		2.281	
Felt depressed	.830		-.789	
Back problems	2.695***		2.367*	
DocDiag. high blood pressure	1.119		1.382	
DocDiag. diabetes	1.131		-2.158	
DocDiag. cancer/tumor	-3.839*		-3.355	
DocDiag. lung problems	-.100		-2.537	
DocDiag. heart attack	4.988***		2.852	
DocDiag. stroke	3.363		3.493	
DocDiag. psych. problem	3.746*		.975	
DocDiag. arthritis/rheumatism	3.199***		1.876	
Change in high blood pressure	-.134		-2.341	
Change in diabetes	9.029**		12.224**	
Change in cancer/tumor	5.577		4.152	
Change in lung problems	4.944		4.612	
Change in heart attack	-6.233*		-2.404	
Change in stroke	12.075*		20.953	
Change in psych. problem	-6.697		-3.675	
Change in arthritis/rheumatism	-1.176		-2.153	
Total houshold income	1.21e-06		2.74e-07	
Total houshold income2	-1.26e-13*		-3.23e-14	
Individual earnings	-.00003		1.00e-05	
Individual earnings2	3.66e-11*		1.24e-13	
Out of pocked health expend.	-7.58e-06		.00006	
Total medical expenditure	.00004**		.00002	
Employed	2.507*		4.304**	
Job requires physical effort	-.443		-2.941**	
Age	.501***		.132	
Male	-2.603***		-.026	
Education > 12 years	.689		.056	
Living with partner	1.357		4.168	
Ever smoked	-.975		.809	
Constant	3.977	27.477***	-6.225	2.037
Number of observations	4274	4930	4274	4494
R^2	.179	.1	.	.
F-statistic	.	453.017	.	164.742

Table 26: Tests for Weak and Strong Rationality using a Linear Probability Model. We only use data from year 1994 and the ExpWorkLimProblem from 1996.

$$\boxed{\text{ExpHealthProblems}_{i,t+1} = \alpha_i + \beta \times ExpHealthProblems_{it} + \Omega_{it}\gamma + \epsilon_{it}}$$

	strongRE	weakRE	IV-strongRE	IV-weakRE
	(1)	(2)	(3)	(4)
ExpHealthProblem	.075***	.207***	1.113***	1.001***
rExpLive to 75	-.081***			
rExpLive to 85	-.065***			
Smokes now	1.560			
Vigorous physical activity	-.153			
Mother alive	1.050			
Father alive	-2.493***			
Health status: Very Good	1.258		-2.717*	
Health status: Good	4.691***		-3.125	
Health status: Fair	7.800***		-3.818	
Health status: Poor	18.273***		-.791	
Diff. walking accross room	-3.277		-11.350	
Body mass index	.053		.009	
Diff. walking 1 block	1.381		-2.960	
Diff. pushing large objects	1.483		-1.320	
Diff. sitting 2 hours	2.159**		1.098	
Diff. using the phone	-3.120		-4.655	
Diff. using money	3.790		9.016***	
Diff. climbing stairs	.496		-4.170***	
Diff. lifting 10 pounds	2.066		-2.646	
Felt depressed	-.175		-1.896	
Back problems	1.742**		1.938*	
DocDiag. high blood pressure	1.381		1.264	
DocDiag. diabetes	.561		-5.293**	
DocDiag. cancer/tumor	-1.349		1.177	
DocDiag. lung problems	1.426		-1.615	
DocDiag. heart attack	4.339***		1.522	
DocDiag. stroke	.115		-1.797	
DocDiag. psych. problem	5.374***		2.740	
DocDiag. arthritis/rheumatism	4.216***		1.100	
Change in high blood pressure	.730		.912	
Change in diabetes	5.010		11.248**	
Change in cancer/tumor	2.975		.117	
Change in lung problems	2.123		6.502	
Change in heart attack	-1.152		3.185	
Change in stroke	4.990		10.917	
Change in psych. problem	-5.219		-4.990	
Change in arthritis/rheumatism	-3.153**		-3.203	
Total houshold income	5.64e-07		1.68e-07	
Total houshold income2	-6.46e-14		-6.67e-14	
Individual earnings	-.00004***		-.00002	
Individual earnings2	4.24e-11**		2.57e-11	
Out of pocked health expend.	-.00003		-.00004	
Total medical expenditure	.00003**		.00003	
Employed	1.978*		5.381***	
Job requires physical effort	.292		-1.351	
Age	.549***		-.107	
Male	-2.716***		-.033	
Education > 12 years	.805		.163	
Living with partner	-1.584		-1.485	
Ever smoked	-1.169		.748	
Constant	6.816	31.634***	2.820	1.191
Number of observations	7411	24465	7411	18124
R^2				
F-statistic				

Table 27: Tests for Weak and Strong Rationality using a Linear Probability Model. We use entire Panel 1992-2002.

Mincer-Type Regression: Log(income$_{it}$) $= \beta \times WorkLimHealthProblems_{it} + x'_{it}\gamma + \epsilon_{it}$	poolOLS	poolOLSr	OLSpanel	Betweeniid	FE-OLS	RE-OLS
	(1)	(2)	(3)	(4)	(5)	(6)
HealthProblems	-.231	-.231	-.231	-.309	-.084	-.162
	(.025)***	(.026)***	(.029)***	(.043)***	(.030)***	(.025)***
Employed	.431	.431	.431	.467	.365	.401
	(.031)***	(.041)***	(.046)***	(.056)***	(.036)***	(.031)***
Job requires physical effort	-.134	-.134	-.134	-.159	-.033	-.094
	(.017)***	(.017)***	(.021)***	(.028)***	(.027)	(.020)***
Hours worked	.048	.048	.048	.058	.026	.039
	(.002)***	(.003)***	(.003)***	(.003)***	(.002)***	(.002)***
Squared hours worked	-.0004	-.0004	-.0004	-.0005	-.0002	-.0003
	(.00002)***	(.00004)***	(.00004)***	(.00003)***	(.00002)***	(.00002)***
Years worked	.041	.041	.041	.041	.119	.041
	(.003)***	(.003)***	(.004)***	(.004)***	(.018)***	(.004)***
Years worked2	-.0006	-.0006	-.0006	-.0006	-.0005	-.0005
	(.00005)***	(.00006)***	(.00007)***	(.00007)***	(.0001)***	(.00006)***
Age	.038	.038	.038	.003	.044	.069
	(.013)***	(.016)**	(.019)**	(.018)	(.032)	(.015)***
Age2	-.0004	-.0004	-.0004	-.0001	-.0009	-.0006
	(.0001)***	(.0001)**	(.0002)**	(.0002)	(.0002)***	(.0001)***
Male	-.327	-.327	-.327	-.362		-.337
	(.018)***	(.018)***	(.023)***	(.028)***		(.027)***
Education > 12 years	.579	.579	.579	.450		.505
	(.141)***	(.155)***	(.187)***	(.218)**		(.171)***
Age*Education>12	-.005	-.005	-.005	-.003	-.002	-.003
	(.002)**	(.003)*	(.003)	(.004)	(.005)	(.003)
Male*Education>12	.042	.042	.042	.028		.046
	(.021)**	(.020)**	(.026)	(.032)		(.031)
Female*Education>12	.066	.066	.066	.074		.080
	(.020)***	(.020)***	(.025)***	(.031)**		(.030)***
Constant	7.150	7.150	7.150	7.479	6.119	5.656
	(.921)***	(.422)***	(.502)***	(84070.950)	(1.171)***	(.452)***
Number of observations	13583	13583	13583	13583	13583	13583
R^2	.377	.377	.377	.457	.078	
F-statistic	210.096	.	.	96.017	21.698	

Table 28: Linear Panel Wave (1-6): Dependent Variable is log(Income)

Mincer-Type Regression: Log(wage$_{it}$) $= \beta \times WorkLimHealthProblems_{it} + x'_{it}\gamma + \epsilon_{it}$	poolOLS	poolOLSr	OLSpanel	Betweeniid	FixEffiid	RandomEiid
	(1)	(2)	(3)	(4)	(5)	(6)
HealthProblems	-.193	-.193	-.193	-.287	-.020	-.089
	(.020)***	(.022)***	(.026)***	(.036)***	(.021)	(.018)***
Employed	.241	.241	.241	.211	.279	.266
	(.024)***	(.032)***	(.038)***	(.045)***	(.025)***	(.022)***
Job requires physical effort	-.107	-.107	-.107	-.135	.00008	-.059
	(.014)***	(.013)***	(.017)***	(.024)***	(.019)	(.015)***
Hours worked	.062	.062	.062	.066	.054	.058
	(.001)***	(.003)***	(.003)***	(.003)***	(.002)***	(.001)***
Squared hours worked	-.0004	-.0004	-.0004	-.0005	-.0004	-.0004
	(.00002)***	(.00004)***	(.00004)***	(.00003)***	(.00002)***	(1.00e-05)***
Years worked	.029	.029	.029	.027	.059	.024
	(.002)***	(.003)***	(.003)***	(.004)***	(.013)***	(.003)***
Years worked2	-.0004	-.0004	-.0004	-.0004	-.0002	-.0003
	(.00004)***	(.00005)***	(.00006)***	(.00006)***	(.00009)**	(.00005)***
Age	.010	.010	.010	-.003	.010	.044
	(.010)	(.011)	(.015)	(.015)	(.022)	(.012)***
Age2	-.0001	-.0001	-.0001	-.00005	-.0002	-.0003
	(.00008)	(.0001)	(.0001)	(.0001)	(.0002)	(.0001)***
Male	-.290	-.290	-.290	-.320		-.264
	(.015)***	(.015)***	(.022)***	(.025)***		(.023)***
Education > 12 years	.080	.080	.080	.033		.203
	(.111)	(.121)	(.160)	(.178)		(.133)
Age*Education>12	.003	.003	.003	.003	.0009	.002
	(.002)	(.002)	(.003)	(.003)	(.003)	(.002)
Male*Education>12	.043	.043	.043	.037		.047
	(.017)**	(.016)***	(.024)*	(.028)		(.027)*
Female*Education>12	.059	.059	.059	.055		.074
	(.017)***	(.016)***	(.024)**	(.027)**		(.027)***
Constant	3.714	3.714	3.714	3.622	3.238	2.191
	(.286)***	(.325)***	(.421)***	(83326.350)	(.620)***	(.648)***
Number of observations	14019	14019	14019	14019	14019	14019
R^2	.495	.495	.495	.534	.264	
F-statistic	350.732	.	.	135.744	94.264	

Table 29: Linear Panel Wave (1-6): Dependent Variable is log(Wage)

Prob(WorkLimHealthProblems$_6 = 1) = \Phi(\beta \times ExpectedWorkLimHealthProblems + \gamma X)$	Probit1	Probit2	Probit3	Probit4	Probit5	Probit6
	(1)	(2)	(3)	(4)	(5)	(6)
ExpHealthProblem	.002 (.001)**	.002 (.0009)**	.002 (.001)**	.002 (.001)*		
ExpLive to 75	.002 (.002)		-.0005 (.001)		-.001 (.0008)	
ExpLive to 85	-.003 (.001)***			-.002 (.0009)**		-.002 (.0008)***
Health status: Very Good	.136 (.083)*	.120 (.079)	.153 (.082)*	.103 (.080)	.165 (.075)**	.138 (.074)*
Health status: Good	.364 (.086)***	.399 (.082)***	.412 (.085)***	.345 (.084)***	.436 (.077)***	.383 (.076)***
Health status: Fair	.448 (.116)***	.457 (.106)***	.463 (.112)***	.411 (.112)***	.508 (.096)***	.430 (.097)***
Health status: Poor	.418 (.208)**	.451 (.181)**	.522 (.190)***	.408 (.204)**	.519 (.133)***	.397 (.138)***
Diff. walking accross room	-.453 (.305)	-.332 (.290)	-.367 (.293)	-.491 (.303)	-.138 (.171)	-.105 (.182)
Diff. walking 1 block	.263 (.104)**	.236 (.095)**	.254 (.097)***	.269 (.103)***	.381 (.079)***	.394 (.082)***
Diff. pushing large objects	.169 (.092)*	.143 (.086)*	.150 (.089)*	.143 (.091)	.209 (.072)***	.226 (.074)***
Diff. sitting 2 hours	.145 (.076)*	.152 (.071)**	.133 (.073)*	.173 (.075)**	.129 (.062)**	.161 (.063)**
Diff. using the phone	.358 (.263)	.227 (.221)	.275 (.244)	.247 (.243)	.192 (.178)	.222 (.192)
Diff. using money	.052 (.137)	-.007 (.128)	-.008 (.133)	.035 (.135)	-.141 (.112)	-.083 (.117)
Diff. climbing stairs	.223 (.069)***	.245 (.065)***	.244 (.067)***	.237 (.068)***	.206 (.059)***	.191 (.059)***
Diff. lifting 10 pounds	.206 (.092)**	.194 (.086)**	.222 (.088)**	.197 (.091)**	.254 (.073)***	.215 (.075)***
Felt depressed	-.004 (.084)	-.006 (.077)	-.0004 (.079)	-.006 (.082)	-.007 (.067)	-.031 (.069)
Back problems	.149 (.064)**	.163 (.060)***	.162 (.062)***	.160 (.062)**	.149 (.054)***	.157 (.054)***
Constant	-3.133 (.481)***	-3.109 (.438)***	-2.947 (.465)***	-3.150 (.455)***	-2.356 (.401)***	-2.653 (.390)***
Number of observations	3380	3707	3538	3496	4399	4342

Table 30: Information Content of WorkLimHealthProblems. Dependent variable is Work Limiting Health Problems in Wave 6. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (these are not reported here).

$$\text{Prob(WorkLimHealthProblems}_6 = 1) = \frac{exp(\beta \times ExpectedWorkLimHealthProblems + X\gamma)}{1+exp(\beta \times ExpectedWorkLimHealthProblems + X\gamma)}$$

	Logit1	Logit2	Logit3	Logit4	Logit5	Logit6
	(1)	(2)	(3)	(4)	(5)	(6)
ExpHealthProblem	.004	.004	.004	.003		
	(.002)**	(.002)**	(.002)**	(.002)*		
ExpLive to 75	.004		-.0007		-.002	
	(.003)		(.002)		(.001)	
ExpLive to 85	-.006			-.004		-.004
	(.002)***			(.002)**		(.001)***
Health status: Very Good	.275	.232	.303	.207	.325	.275
	(.157)*	(.149)	(.156)*	(.152)	(.142)**	(.138)**
Health status: Good	.689	.738	.776	.646	.808	.707
	(.160)***	(.150)***	(.158)***	(.155)***	(.142)***	(.139)***
Health status: Fair	.827	.826	.854	.750	.912	.772
	(.206)***	(.188)***	(.199)***	(.198)***	(.172)***	(.171)***
Health status: Poor	.766	.811	.938	.739	.925	.705
	(.362)**	(.312)***	(.331)***	(.354)**	(.234)***	(.241)***
Diff. walking accross room	-.773	-.568	-.628	-.842	-.180	-.134
	(.525)	(.496)	(.501)	(.520)	(.302)	(.321)
Diff. walking 1 block	.421	.378	.404	.433	.616	.640
	(.174)**	(.159)**	(.163)**	(.172)**	(.133)***	(.138)***
Diff. pushing large objects	.292	.246	.258	.244	.355	.384
	(.158)*	(.146)*	(.151)*	(.155)	(.123)***	(.125)***
Diff. sitting 2 hours	.241	.250	.220	.291	.206	.268
	(.132)*	(.122)**	(.127)*	(.128)**	(.106)*	(.108)**
Diff. using the phone	.605	.352	.441	.401	.325	.402
	(.468)	(.383)	(.430)	(.428)	(.315)	(.339)
Diff. using money	.104	-.004	-.007	.073	-.229	-.118
	(.236)	(.223)	(.230)	(.233)	(.198)	(.205)
Diff. climbing stairs	.396	.432	.432	.418	.359	.329
	(.119)***	(.111)***	(.115)***	(.116)***	(.101)***	(.101)***
Diff. lifting 10 pounds	.337	.320	.366	.322	.423	.359
	(.158)**	(.147)**	(.151)**	(.156)**	(.123)***	(.127)***
Felt depressed	-.017	-.014	-.007	-.022	-.013	-.058
	(.146)	(.133)	(.138)	(.143)	(.116)	(.121)
Back problems	.253	.277	.275	.272	.250	.262
	(.111)**	(.104)***	(.107)**	(.108)**	(.093)***	(.093)***
Constant	-5.586	-5.475	-5.248	-5.546	-4.121	-4.607
	(.868)***	(.779)***	(.833)***	(.813)***	(.715)***	(.693)***
Number of observations	3380	3707	3538	3496	4399	4342

Table 31: Information Content of worklm, dependent variable is Work Limiting Health Problems in Wave 6. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (these are not reported here).

	1 Restricted	2 Restricted	3 Unrestricted
σ_1 :	$0.0157\ (0.0059)$	$0.0229\ (0.0082)$	$0.0241\ (0.0133)$
σ_2 :	$0.0453\ (0.0124)$	$0.0397\ (0.0091)$	$0.0191\ (0.0091)$
Ψ :	1	1	$2.3706\ (0.5119)$
LogLikelihood:	-3956.821	-2953.354	-2881.642
Sample Size:	7001	7001	7001
Data Type:	5th deg. polyn	original	original
Iterations;	146	90	250
Estimation Time:	1.39 hours	0.81 hours	2.11 hours

Table 32: Estimation Results from Maximum Likelihood Estimation

10.6 Appendix F: Figures

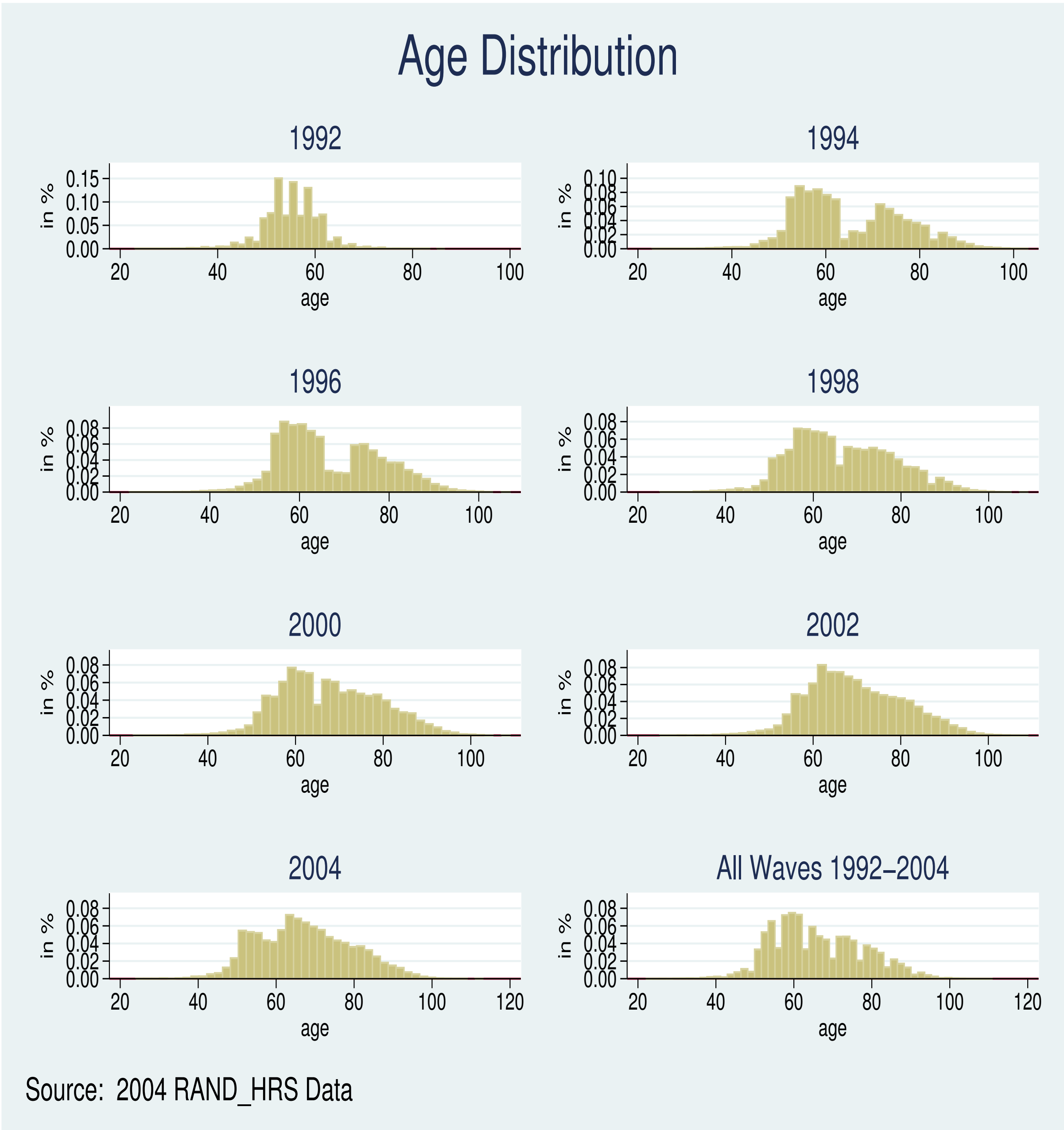


Figure 1: Histograms of Age

Figure 2: Histogram of Expected Work Limiting Health Problems of Males: Age 40-60.

Figure 3: Histogram of Expected Work Limiting Health Problems of Females: Age 40-60.

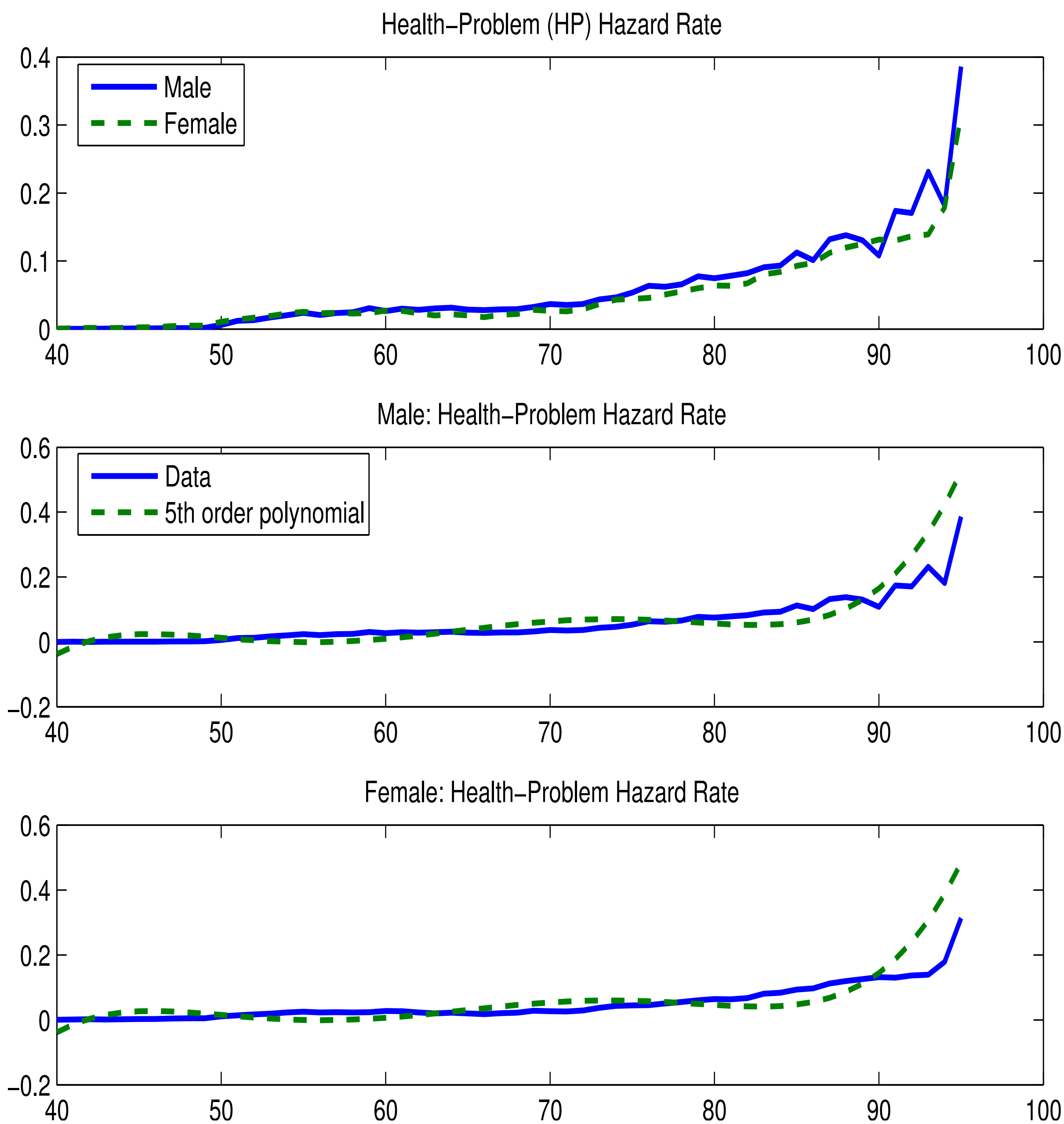


Figure 4: Work Limiting Health Problems Hazard Rate. Original Data from RAND-HRS,Wave 1-6. Fitted function is a 5th order polynomial, fitted with least squares.

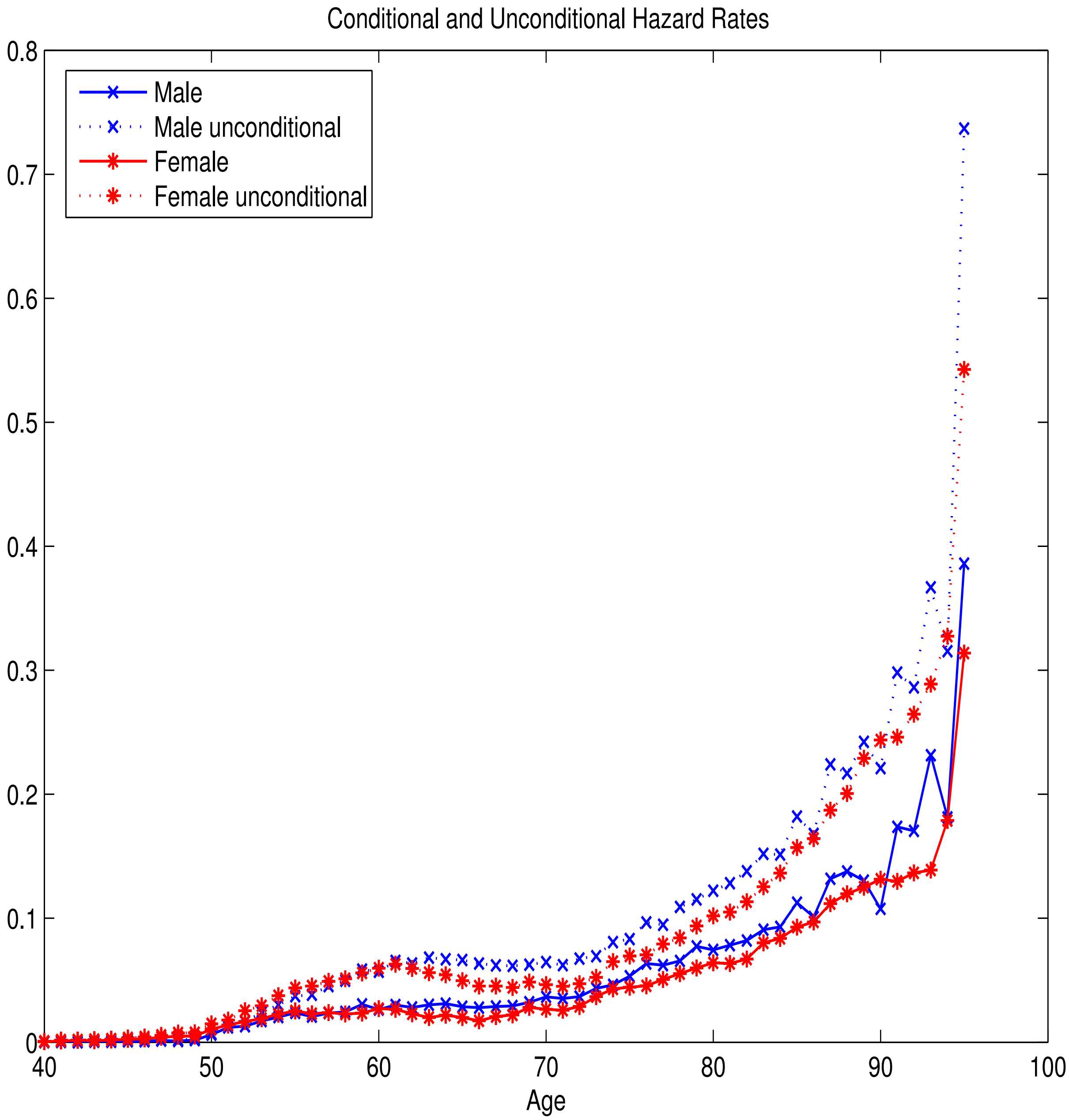


Figure 5: Conditional and Unconditional Hazard Rates of Developling Work Limiting Health Problems. Original Data from RAND-HRS,Wave 1-6.

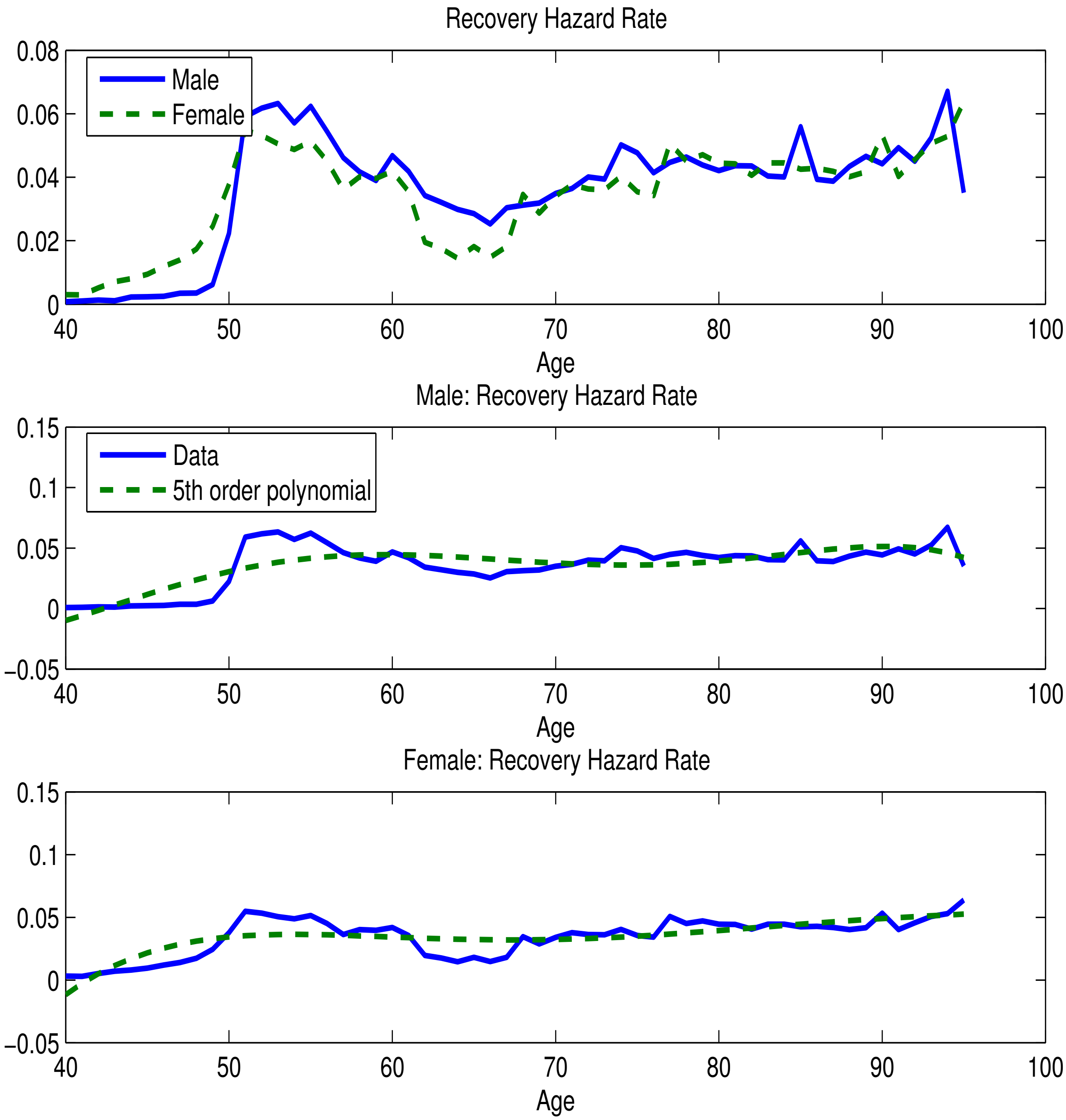


Figure 6: 'Recovery from Work Limiting Health Problems' Hazard Rate. Original Data from RAND-HRS,Wave 1-6. Fitted function is a 5th order polynomial, fitted with least squares.

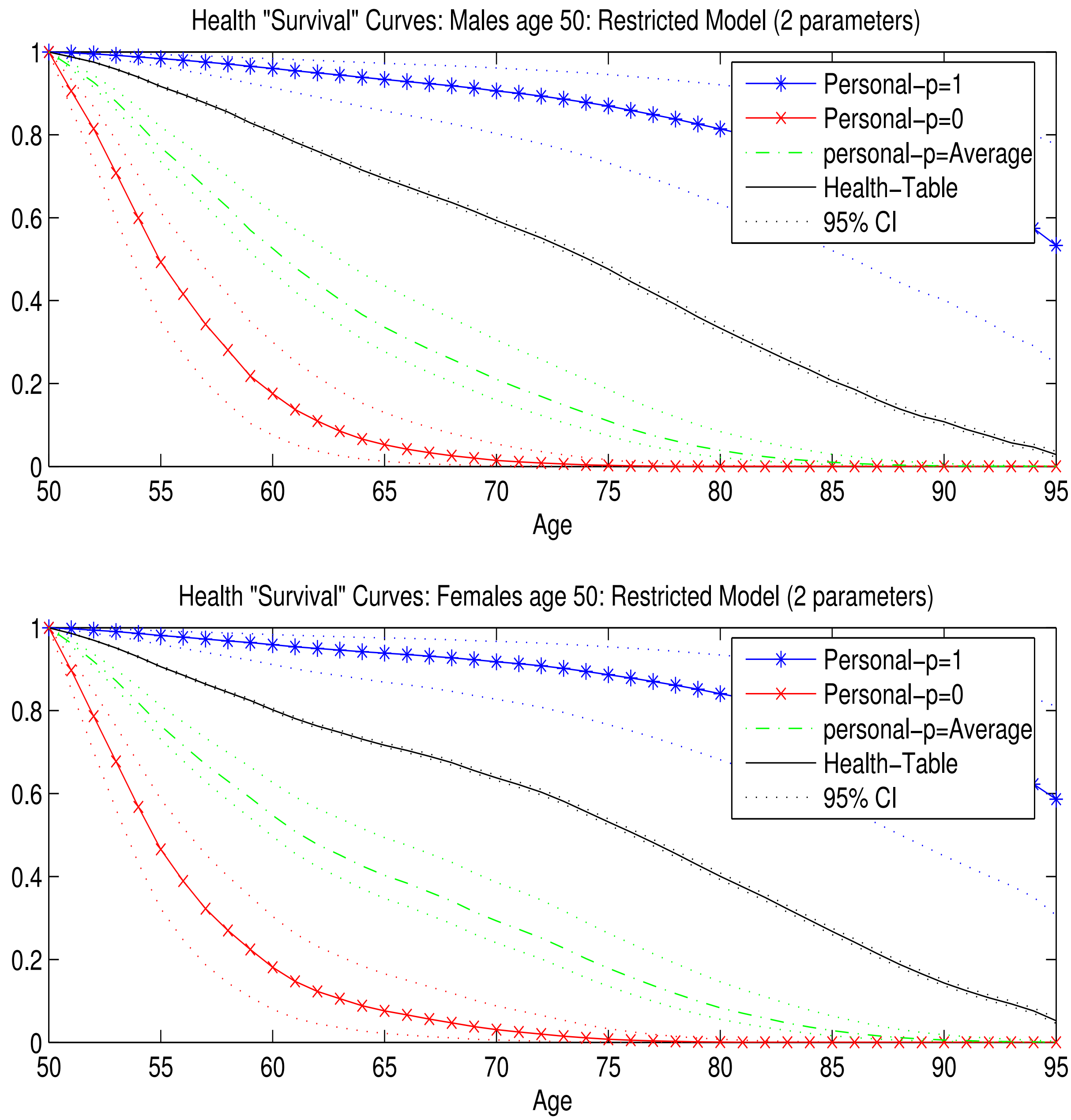


Figure 7: Health "Survival" Probabilites of a 50 Year Old.

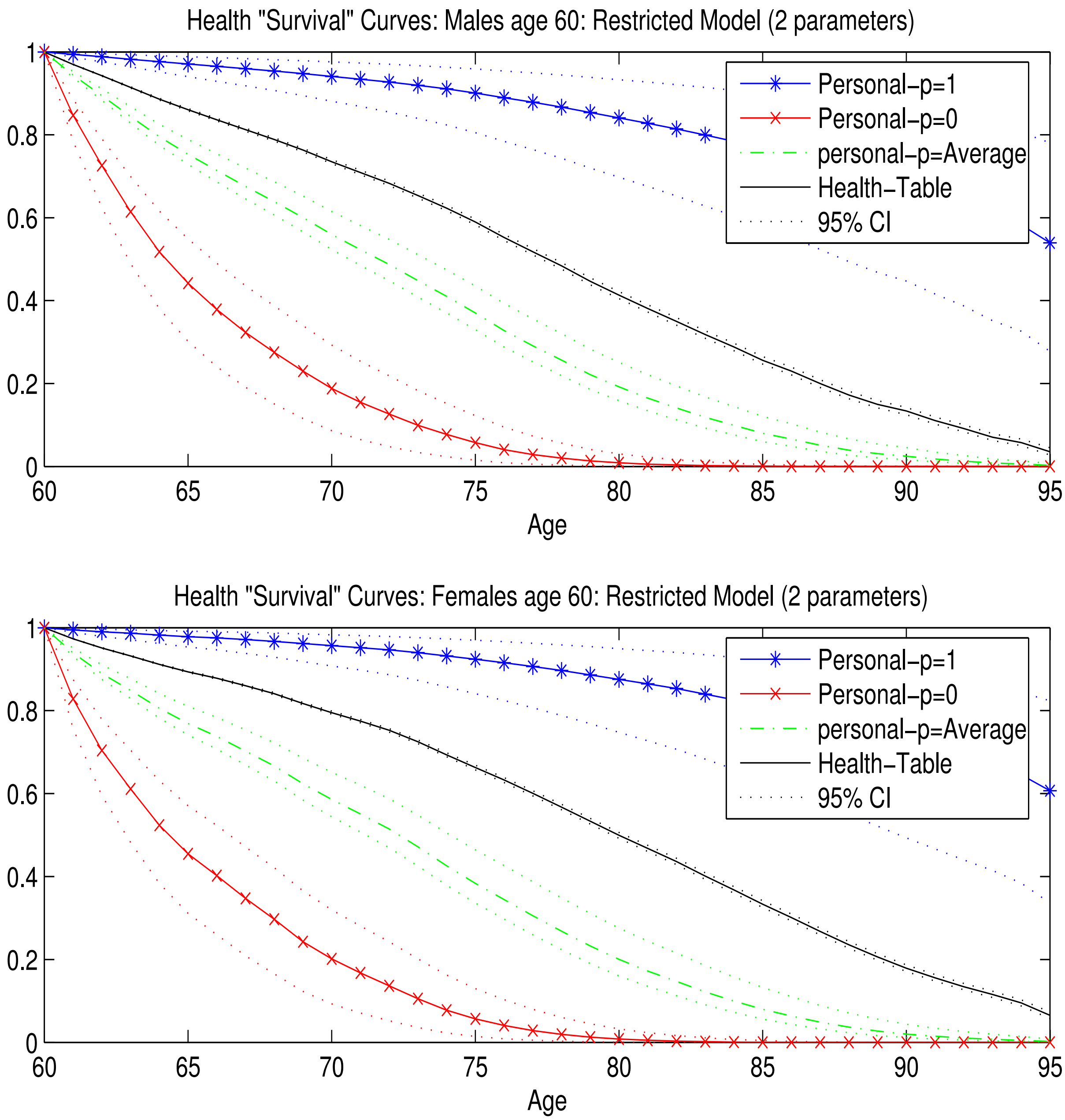
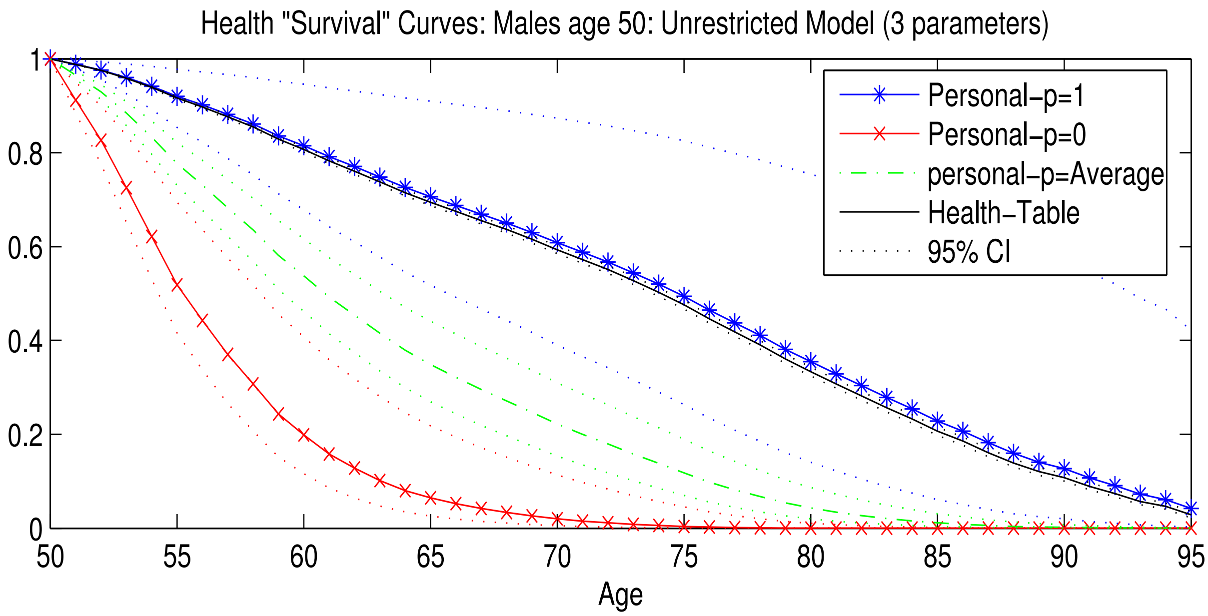


Figure 8: Health "Survival" Probabilites of a 60 Year Old.



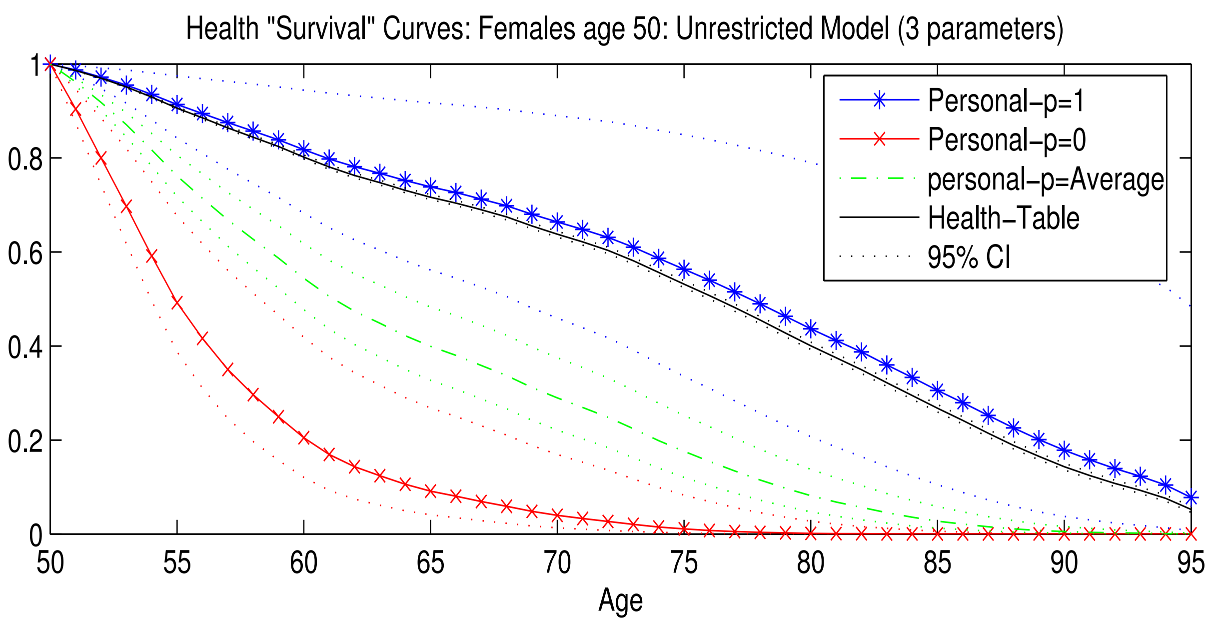


Figure 9: Health "Survival" Probabilites of a 50 Year Old for the Unrestricted Model.

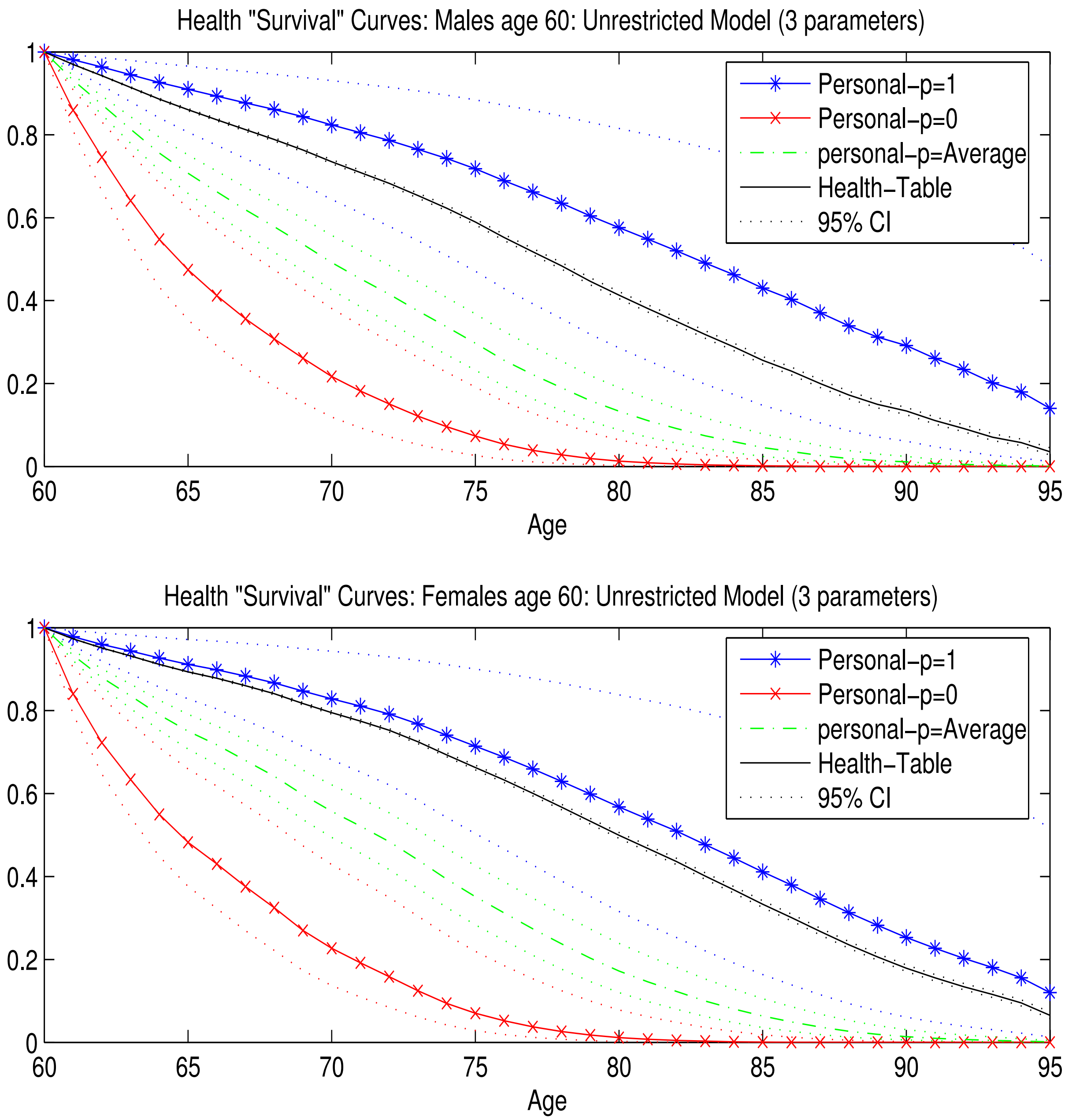


Figure 10: Health "Survival" Probabilites of a 60 Year Old for the Unrestricted Model.

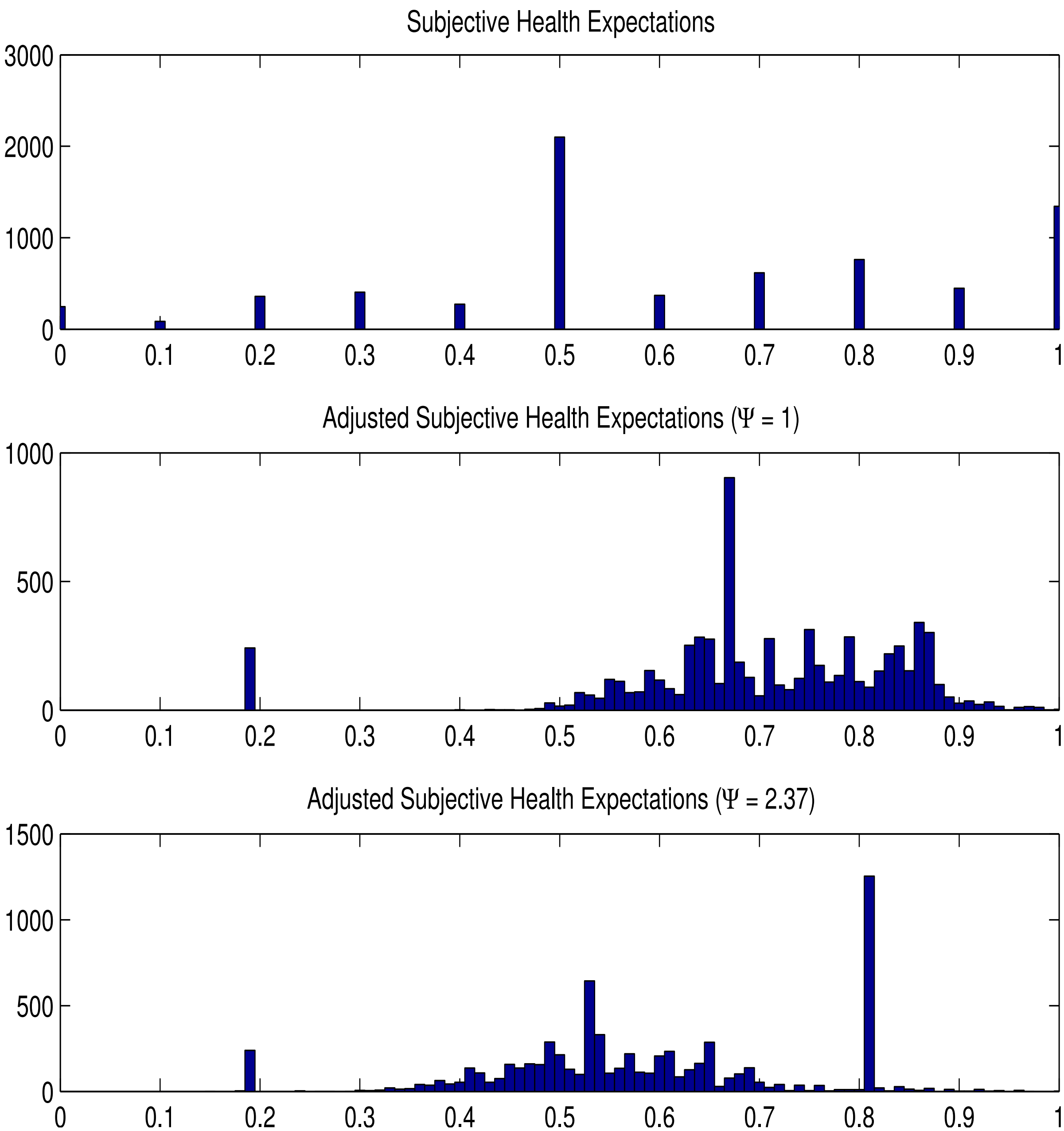


Figure 11: Histograms of Subjective Health Expectations

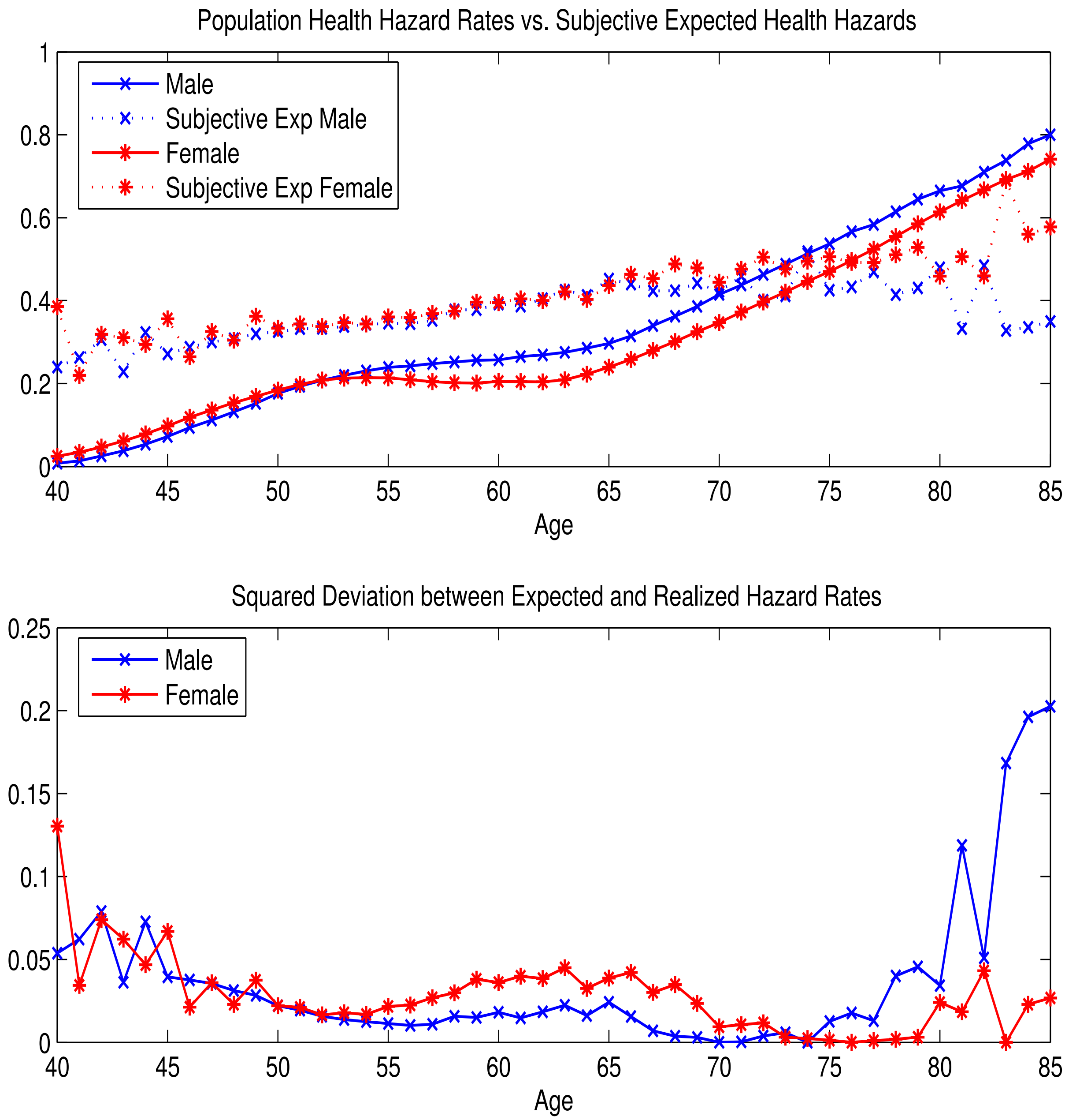


Figure 12: Population Health Hazard (10 year cumulative) vs. Mean Subjective Hazard Rate,Wave 1-6.

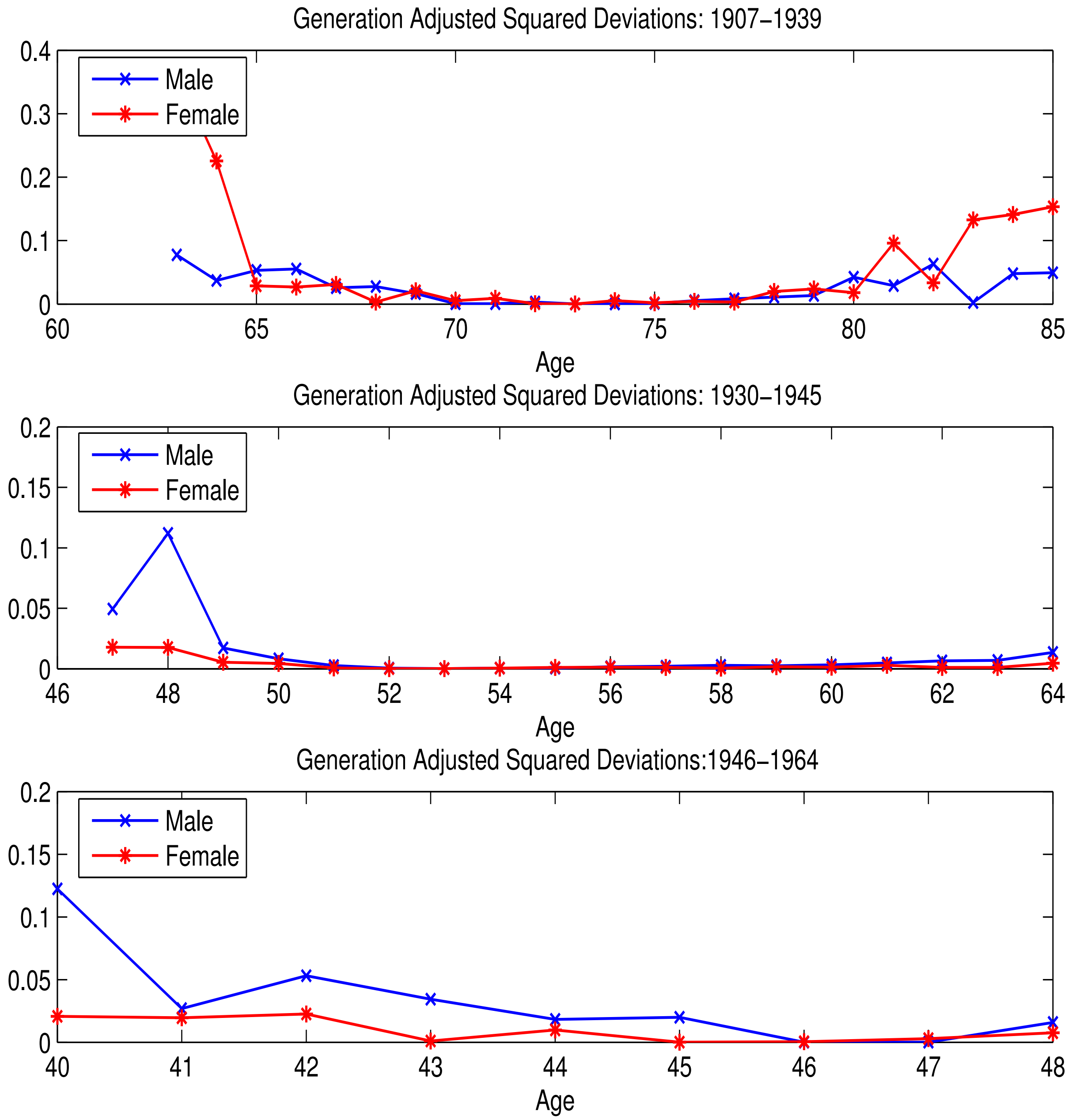


Figure 13: Squared Difference between Population Health Hazard (10 year cumulative) vs. Mean Subjective Hazard Rate by Generation.