University of Michigan Retirement Research Center

Working Paper
WP 2003-037

What to Expect when you are Expecting Rationality: Testing Rational Expectations using Micro Data

Hugo Benítez-Silva and Debra S. Dwyer

MRRC Project #: UM02-11

"What to Expect when you are Expecting Rationality: Testing Rational Expectations using Micro Data"

Hugo Benítez-Silva
SUNY – Stony Brook

Debra S. Dwyer
SUNY – Stony Brook

June 2003

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104

Acknowledgements

This work was supported by a grant from the Social Security Administration through the Michigan Retirement Research Center (Grant # 10-P-98358-5). The opinions and conclusions are solely those of the authors and should not be considered as representing the opinions or policy of the Social Security Administration or any agency of the Federal Government.

What to Expect when you are Expecting Rationality: Testing Rational Expectations using Micro Data

Hugo Benítez-Silva
Debra S. Dwyer

Abstract

This paper tests the Rational Expectations (RE) hypothesis regarding retirement expectations, controlling for sample selection, reporting biases, and unobserved heterogeneity. We find that retirement expectations in the Health and Retirement Study (HRS) are consistent with the RE hypothesis. We also examine how a wide array of factors, such as wealth, income, health insurance, pensions, and health status influence retirement expectations formation using panel data from all available waves of the HRS. We further analyze how new information affects the evolution of retirement expectations and discover that, on average, individuals correctly anticipate most uncertain events when planning their retirement, except for some health conditions and economic factors. Our results have important implications for a wide variety of models in economics that assume rational behavior.

Keywords: Rational Expectations, Retirement Expectations, New Information
JEL classification: D84, J26

Authors' Acknowledgements

The authors are grateful to the Michigan Retirement Research Center (MRRC) and the TIAA-CREF Institute for financial support. An early version of this paper circulated with the title "*Expectations Formation using the Health and Retirement Study.*" We are grateful to Moshe Buchinksy, John Rust, Tom Muench, Jim Hosek, Michael Hurd, Arie Kapteyn, Stan Panis, James P. Smith, Christian Belzil, Bryan Campbell, Nikolay Gospodinov, Jenny Hunt, Peter Arcidiacono, Ahmed Khwaja, Marjorie McElroy, Julia Lynn Coronado, Don Andrews, Valentina Corradi, Albert Marcet, Ernesto Villanueva, Douglas Fore, John Ameriks, Chris Swann, Olivia Mitchell, and Anna Vellvé-Torras for insightful comments and discussions. We also would like to acknowledge excellent research assistance from Huan Ni, Selcuk Eren, and Wayne-Roy Gayle. We have benefited from comments from participants at the Society of Labor Economists Meetings in Baltimore in May 2002, the North American Meetings of the Econometric Society at UCLA in June 2002, the RAND Labor and Population group, the CIREQ Labor and Applied Econometrics Workshop at Concordia University, the Applied Microeconomics Seminar at Duke University, and the Labor and Public Economics Seminar at UPF. All remaining errors are the authors'.

1. Introduction

The Health and Retirement Study (HRS) provides us with the rare opportunity to directly analyze the evolution of retirement expectations and test the Rational Expectations hypothesis which is without a doubt one of the more used and useful hypothesis in economics. Due to its extensive use, it is surprising how widely the definition of the term Rational Expectations varies.[1] Our use of the term and the approach to testing the hypothesis will be very close to and consistent with the views expressed by the precursors of this powerful assumption. We will essentially maintain that *agents' subjective beliefs about the evolution of a set of variables of interest coincide with the objectively measurable population probability measure*. This is in line with the original characterization of Muth (1961) and Lucas (1972).[2] The main difference with this seminal work, however, is that instead of concentrating on forecasts of market level variables, we will focus on how individuals form expectations over an array of micro variables that affect them directly, or are in part under their control, for instance their retirement age, their health, their employment, their wealth, their income, etc.

It is important to distinguish that while in macroeconomic theory the RE assumption is understood mainly as an equilibrium concept,[3] in microeconomic applications the concept is used as a synonym of individual rationality or an efficient use of information with regard to individual level variables. The latter implies that somehow the economy is in equilibrium. This RE assumption at the micro level underlies a majority of the research in applied fields, and it is at the forefront of most work in dynamic models of individual behavior. This is primarily because most household and individual level data sources are rich in micro variables and it is the responsibility of the researcher to try to control for the macroeconomic environment in which those decisions are made.

The debate over whether testing rational expectations is a worthwhile enterprise goes back more than two decades. Prescott (1977, p.30) expressed a strong opinion, one that Young and Darity (2001) argue he still maintains:

---

[1] A similar statement could have been made (and might still be) about the term 'rationality', as discussed, for example in Schuetz (1943). However, economics has come a long way since then in formalizing our methodological approach to economic analysis, with the characterization of rationality playing a central role in that leap.

[2] For a survey of the early contributions see the special issue in the *Journal of Money, Credit and Banking* edited by McCallum (1980). For a more recent discussion see Sargent (1995).

[3] Thanks largely to Lucas' seminal contributions, where expectations affect the stochastic evolution of the economy and this evolution in turn affects expectations formation.

"Like utility, expectations are not observed, and surveys cannot be used to test the rational expectations hypothesis. One can only test if some theory, whether it incorporates rational expectations, or, for that matter, irrational expectations, is or is not consistent with observations."

Part of our work in this paper tries to tackle that skepticism, and make the case for retirement expectations from the HRS as being a measure of *something* (Young and Darity 2001, quote Prescott as saying that responses to expectations questions should not be treated as measurements of *anything*), in this case retirement itself.

Once we make that case we will be ready to go further and follow the *mandate* from Simon (1979, p.505),

"It is too early to render final verdict on the rational expectations theory. The issue will ultimately be decided, as all scientific debates should be, by a gradual winnowing of the empirical evidence, and that winnowing process has just begun."

Tobin (1980, p.29),

"Economists, all of us, should pay more attention to actual data on expectations and how they are formed and less to our own assumptions about what they are and how they are or should be formed."

And Lovell (1986, p.111),

"My own view is that the appropriate realm for empirical research should not be demarcated in terms of a dichotomy between assumptions and predictions—I think that direct testing of the rational expectations hypothesis is an appropriate and worthwhile activity…a theory that is said to be based on micro foundations should survive empirical testing at the level of the individual decision making unit."

We test the Rational Expectations Hypothesis using data on retirement expectations, and we explore how these expectations are influenced by individual and household factors such as wealth, income, health insurance, pensions, and health status. We also study how and whether new information affects plans over time. Once we confirm rationality, we analyze the role of new information to better understand expectation formation.

The objective of testing the RE Hypothesis also goes back a couple of decades, as the debate above suggests. It begins in the context of the life cycle permanent income hypothesis in a stream of literature that starts with the seminal contribution of Hall (1978), and then compares forecasts of market variables with realizations like in Figlewski and Watchtel (1981 and 1983), Kimball Dietrich and Joines (1983), de Leeuw and McKelvey (1981 and 1984), Gramlich (1983), and more recently Davies and Lahiri (1999), and Christiansen (2003). Finally, work by Leonard (1982) analyzes wage expectations of employers, and Fair (1993) analyzes the question in the context of large macroeconomic models. In all these cases the concern is with market level variables, and the evidence in these and many other studies is mixed. We propose a slightly different approach in line with Bernheim's (1990) work, which we believe is likely to bring back to center stage the debate about whether we should be testing one of our key working hypotheses, in particular the RE hypothesis, now that large data sets provide us with the opportunity to explore expectations over micro variables in more detail than ever before.[4]

Panel data available through the HRS allow us to observe people's plans as they approach retirement, and allow us to control for unobserved heterogeneity regarding retirement planning by exploiting the longitudinal nature of the data. We take advantage of the richness of the information we are using, not only in evaluating factors that influence retirement plans over time, but also in analyzing how consistent those plans are and how new information influences those plans. We want to understand which factors matter when planning for retirement, including how people make use of existing and new information to form and update their expectations. In particular, how do people form expectations over uncertain events and how should uncertainty be modeled?

If our tests reject the Rational Expectations (RE) hypothesis, two very different, but nonetheless connected, interpretations are possible.[5] First, we could conclude that models of rational behavior expect too much of individuals, forcing us to abandon the "full rationality hypothesis" that economic agents behave "as if" they were making a large number of computations implied by the theory. One possible alternative to the fully rational model could be an adaptive learning model which introduces a form of bounded rationality, in which individuals use standard econometric techniques to make and adjust their forecasts of relevant variables, with RE emerging as an equilibrium of this trial and error process (see, for example, Pesaran 1987, and Evans and Honkapohja 2001 for illuminating presentations of these type of models). Second, we could conclude that reality is much more complex than even most dynamic models assume, with individuals forming expectations (maybe rational ones) not over a fixed probability distribution of uncertain events, but over a family of distributions for each source of uncertainty.

[4] For example the NLSY79 asked respondents in some survey years their expected number of years of education. Dominitz and Manski (1996 and 1997) analyze the use of data on wage expectations by households and young adults in two different surveys.

[5] A third, less exciting alternative is that one of the assumptions we are jointly testing along with rationality is rejected. We will later make clear the nature of these additional assumptions, which happen to essentially coincide with those also jointly tested along with RE by previous researchers.

This involves individuals learning over time about the characteristics of these distributions and updating their priors as new information comes along.

The first conclusion would be a set back for research using dynamic modeling since it would put into question an attractive and central tool. The second, would mean that we need more realistic dynamic models, which are likely to be more complex, but also more attentive to details of the process of expectations formation by individuals. In essence the latter would be equivalent to revisiting Muth's (1961, p. 316) point that "*dynamic economic models do not assume enough rationality*."

Finally, if our tests do not reject the rational expectations hypothesis, we can at least continue to rely on that rationality, and the strategies used to model it, as a good first approximation to behavior by economic decision makers. Furthermore, it would then be reasonable to use some of these variables in modeling complex economic situations as van der Klaauw (2000) does.

We discuss the literature and our contributions in Section 2. The conceptual model and the econometric specifications are presented in Section 3. Section 4 provides information about the data used in the empirical analysis, and Section 5 reports our main findings. We conclude in Section 6.

2. Background

Most of the retirement literature, using both reduced form and structural models, focuses on the retirement decision or "outcome", with an emphasis on the role of health, social insurance, and economic status of agents in the household. We take a different approach and analyze how people's retirement expectations evolve over time. We examine the importance of financial incentives, health status, and health insurance status on expectations. Understanding how these and other factors affect expectations would give us a better sense of the importance of these incentives on behavior.

Comparing Expectations to Outcomes

Much of the work on expectations of retirement focuses on comparing those expectations against outcomes (Bernheim 1988, 1989; Dwyer and Hu 1999; Forni 2002, Dwyer 2002). Deviations are explained by unanticipated changes to relevant factors. These types of analyses concentrate on how precise expectations are in terms of their relationship to the outcome and not on how expectations are formed. Only Bernheim (1989) provides tests in the tradition of the RE literature, but due to data limitations the results are mixed.

Bernheim (1988 and 1989) explores the connection between expectations and outcomes and finds that although people do not use all the information available to them, they do form rational plans and stick to them. The main finding is that people report likely outcomes instead of mean dates given a probability distribution.[6] He uses the Retirement History Survey (RHS), which represents a cohort retiring in the seventies. The data were not as rich in their ability to tackle retirement expectations, particularly in health status.

More recent work by Dwyer and Hu (1999) and Dwyer (2002) also examines this question in a static life cycle framework comparing baseline plans to outcomes. Dwyer and Hu (1999), using only the first two waves of the HRS, find that people who retire early as planned tend to be able to afford to do so and they are more likely to have sources other than Social Security benefits to pay for additional leisure. The authors conclude that health is a very important predictor of retirement plans, while economic factors become more influential in determining actual retirement.[7] Dwyer (2002), also using the HRS, compares baseline expectations (ages 51-61) of retirement to outcomes by wave 4 (ages 57-67) when a substantial sub-sample has reached its expected retirement age. She finds that the majority of the HRS sample follows through on plans for retirement if you include partial retirement in the definition. Health and socio-economic factors, however, still play a role in explaining changes to plans.

This literature ignores all of the information available between the initial baseline plan and the ultimate outcome at the final stage. It is important to account for the adjustment process which occurs in between. In the first round of interviews, not everyone has thought about retirement, and we only observe plans for those who have which leads to heterogeneity concerns.[8] There is a sample selection problem that is exacerbated by focusing on only those who thought about it early on in the process. In this work we control for this sample selection, and we utilize

[6] This is the mode of the distribution. As Bernheim emphasizes this is not a problem as long as the distribution is symmetric and singled peak. It should also be the case that the sample average of the reported most likely outcomes coincides with the mean of the distribution.
[7] This evidence might indicate that the market incompleteness inherent to the Social Security system (the inability to borrow against future benefits) has the effect of making current economic measures poor predictors of future economic variables in reduced form models.
[8] Some people report expected ages of retirement even though they admit to not having thought about it much. A majority of those who have not thought about it do not report any expected age. Many say they will likely 'never' retire if they do report something. Prior work uses any age with no adjustment for the amount of planning that went into it. That work also uses proxies of benefit take-up for expected age if it is missing.

the five waves of data and examine how expectations are formed and how new information affects expectations over time once we have tested the RE hypothesis.

Forni (2002) uses the first two waves of the HRS to examine how financial shocks influence deviations to plans. He uses self-reported financial shocks which are quite noisy and do not perform well in the models. He concludes that the data are consistent with Bernheim's finding that reported expectations measure likely retirement outcomes rather than the mean of the expected retirement age distribution.

Recent work by Coronado and Perozek (2001) examines the effect of unanticipated changes in wealth on retirement and finds that households were more likely to retire early if they held more corporate equity immediately prior to the 1990s' bull market. They introduce uncertainty to financial factors and examine the retirement outcome. Hurd and Retti (2001) tackle a similar research question using the self-reported probability of working beyond age 62, but find no evidence in changes in this variable with the wealth increases in the stock market in the mid to late 1990s. We, too, examine a related question regarding uncertain factors influencing retirement but we focus on retirement expectations over time. Lusardi (1999) focuses on the importance of information costs in the retirement decision and savings, which acknowledges uncertainty in a different way from the present project.

Expectations Formation

There has been little work done on expectation formation. Bernheim (1990) focuses on expectations formation to test individuals' rationality. He cannot reject the hypothesis of strong rationality, meaning that only new information affects individual changes of expectations regarding Social Security receiving benefit levels.[9]

We build on Bernheim's model of expectations formation, and on the tradition of Muth (1961) and the rational expectations revolution, to argue that testing whether past retirement expectations are a sufficient statistic for current retirement expectations is indeed a test of the RE hypothesis. We use the full longitudinal HRS, and we measure retirement expectations differently and with different instruments to deal with endogeneity and reporting biases problems. We also use panel data techniques and control for sample selection.

Dwyer and Mitchell (1999) study expectation formation using only wave 1 of the HRS. They incorporate health into models of labor supply, while dealing with the potential endogeneity of health. They use expectations

[9] More recently, Dominitz, Manski, and Heinz (2001) using data from the Survey of Economic Expectations have analyzed the connection between Social Security Expectations and retirement savings, but do not discuss individual rationality.

of retirement as a proxy for retirement. The assumption is that the expectation is equal to the outcome. Our analysis can be considered a test of their maintained hypothesis.

Dynamic Models

The research discussed so far uses static life cycle models of the retirement decision and applies it to planning for retirement to test some level of rationality. Since we are also interested, however, in pursuing this question later in a dynamic framework, we want to test the Rational Expectations hypothesis to understand how to appropriately incorporate uncertainty in dynamic models. The rational expectations assumption is a cornerstone of most dynamic utility maximization models under uncertainty.

Recent literature on dynamic modeling has been able to solve rich dynamic life cycle models of retirement, allowing for uncertainty over future wages, future investment returns, health outcomes, and even spouse's variables. Dynamic models to date have not focused on expectation formation with respect to the variables considered stochastic. Models to date usually assume the variables follow a certain distribution, which is then integrated out using numerical or mathematical methods. We propose to take a closer look at how expectations over uncertain outcomes are formed, and how they influence current decisions. If we are unaware of how these expectations are formed, and whether they are realized, we might be missing an important dimension of the decision making process by rational agents, potentially leading to less efficient policy recommendations. In this work, we test the Rational Expectations hypothesis before assuming it in the next stage of this and future research. Rejecting the RE hypothesis might cast serious doubts on a key assumption of dynamic models, maybe suggesting that either simpler models of individual behavior are to be used, or on the contrary, more realistic dynamic models that account for learning about the probability distributions of future events might be worth considering.

Our results have important implications for the growing literature on dynamic modeling, and models that rely on some form of a rationality assumption to solve for, simulate, and eventually estimate the parameters of an economic model of behavior.

3. A Model of Expectations Formation

Define

$$X_t^e = E\langle X \,|\, \Omega_t \rangle, \qquad (1)$$

where the variable X, retirement age, is forecasted, the full information set at time *t* is represented by Ω_t, and *E* is the expectations operator. This is the most commonly used representation of the Rational Expectations hypothesis, where we take as the rational expectation of a variable its conditional mathematical expectation (see for example Sargent and Wallace 1976). This guarantees that errors in expectations will be uncorrelated with the set of variables known at time *t*.

Variables included in the vector representing the information set Ω, come from standard life cycle models of retirement behavior (see Lumsdaine and Mitchell 1999 for a survey). It is well established that the factors that influence retirement include, for example, health, health insurance, and socio-economic status. Our first efforts focus on examining the role of these factors on retirement expectations. A logical hypothesis at this point is that the factors that influence retirement also influence retirement expectations in a similar way.

Using the law of iterated expectations and assuming that the new information was correctly forecasted by agents, from equation (1) we get:

$$E\langle X_{t+1}^e \,|\, \Omega_t \rangle = E[E\langle X \,|\, \Omega_t, \omega_{t+1} \rangle \,|\, \Omega_t] = E\langle X \,|\, \Omega_t \rangle = X_t^e, \qquad (2)$$

where ω_{t+1} represents information that comes available between periods *t* and *t+1*. Notice that it is essential here to assume that new information (its conditional distribution, not just its mean) is correctly forecasted. Without this additional assumption expression (2) would not be correct. We are going to test this assumption jointly with the more standard RE hypothesis, once we also assume linearity of the process presented in (2). Notice that the assumption of correct forecasting is in essence no different from the assumption in the early RE literature; namely that forecast errors are normally distributed with mean zero, in a specification that regresses outcomes of a particular market variable on its expectations and a constant.

Then from (2) we can write the evolution of expectations through time as

$$X_{t+1}^e = X_t^e + \eta_{t+1}, \qquad (3)$$

where

$$\eta_{t+1} = X^e_{t+1} - E[X^e_{t+1} \mid \Omega_t],$$

and therefore $E(\eta_{t+1}|\Omega_t)=0$. Notice that η_{t+1} is a function of the new information received since period *t*, ω_{t+1},

From this characterization of the evolution of expectations, as in Bernheim (1990), we can test the Rational Expectations (RE) hypothesis with the following regression:

$$X^e_{t+1,i} = \alpha + \beta X^e_{t,i} + \gamma\ \Omega_{t,i} + \varepsilon_{t,i}\,, \qquad (4)$$

where α is a constant, and γ is a vector of parameters that estimate the effect of information in period *t* on period's *t+1* expectations. The RE hypothesis implies that $\alpha=\gamma=0$, and $\beta=1$. A weak RE test, in the terminology of Lovell (1986) and Bernheim (1990), assumes that γ is equal to a vector of zeros, and tests for $\alpha=0$ and $\beta=1$—in other words it tests whether expectations follow a random walk. The strong RE test is less restrictive and also tests for $\gamma=0$. Bernheim (1990) cannot reject these hypotheses in a model where agents are forming expectations regarding the level of Social Security benefits they will receive, once he controls for measurement error in the self-reports using instrumental variables.[10] In the next sections we will test these hypotheses exploiting the longitudinal nature of the HRS, while controlling for measurement error and sample selection.[11]

We are also interested in the role of new information in expectations formation. We want to know whether new information affects the evolution of expectations, and whether this new information is to some degree anticipated.

Even assuming we cannot reject the prediction of the RE hypothesis, that in (4) $\beta=1$, a characterization of the role of new information in the evolution of retirement expectations implies that individuals integrate old and new information when updating their expectations, attaching weights to each of these sources. Assuming the linearity of the function that individuals use to process information regarding the variables of interest, and further assuming that the elements of the information set Ω are jointly normally distributed we can then write

[10] He uses other expectations as instruments, like expected retirement age, that may be equally plagued with measurement error. He does not test the instruments as we do. We will discuss our instruments and tests in the results section. The possible major role of measurement error in RE tests was already emphasized by Lovell (1986).

$$X^e_{t+1i} = X^e_{t,i} + \gamma\ (\omega_{t+1} - E[\omega_{t+1} \mid \Omega_t]), \qquad (5)$$

where the term in parenthesis is the difference between the observed new information and the expectation individuals had of this information as of last period. If we had all these elements we could then estimate (5) and would expect a coefficient of one in front of the retirement expectation as of time *t*, and the coefficients in γ could be interpreted as the effect on changes in expectation of the unanticipated components of new information.[12] However, we do not observe the second element in the brackets, and although it could be estimated by making further assumptions about the relationship between new and current information, this is likely to be a very noisy procedure. Instead, we estimate the following equation

$$X^e_{t+1i} = \alpha + \beta\ X^e_{t,i} + \gamma\ \omega_{t+1,i} + \varepsilon_{t,i}, \qquad (6)$$

where there is no reason to believe that β would be equal to 1, since the unobserved expectation term from (5) would enter into the error term, leading to possible biases in the coefficients of interest, which in part can be treated as a problem of measurement error and attenuated using IV techniques. Also, the estimates of the vector γ, are interpreted as the effects that changes in specific factors in Ω have on expectation formation. A nonzero coefficient on a given change in a member of the information set can be interpreted as information not perfectly anticipated, and therefore not embedded in period's *t* expectations.

Econometric Specifications

Before we can give empirical content to the model above we are faced with a number of issues that need to be discussed. First, we need to appropriately control for observed heterogeneity. This is not trivial when we consider omitted variables as well as potential reporting biases. How to incorporate relevant factors that are imperfectly observed has its own set of issues. In addition, respondents make decisions based on many factors that may not be observable to the econometrician. Given these measurement error and omitted variable bias issues, unobserved heterogeneity becomes a concern. We want to take into account the unobserved heterogeneity potentially present in our characterization of the econometric model. If we do not control for the unobserved

[11] Rosen (1990) in his commentary on Bernheim's work acknowledges the potential for selection bias in responses to survey questions about expectations. He also emphasizes the importance of revisiting the issue of rational planning with newer and better data.

[12] Depending on the covariance structure of model used by individuals to process information, these coefficients could be time dependent. This could potentially affect the strategy to estimate these coefficients. In the sequel we will assume they are time invariant. We thank Tom Muench for helping us understand this point.

components we will be confounding partial and total effects of our variables of interest. Panel data sets allow us to model explicitly how those unobserved components enter the econometric specification, and we can choose to include them as a fixed effect or as a random variable, and test the different specifications.[13]

Related to this is sample selection. Not only do people differ in their far-sightedness, but also in their ability to process information. For this reason we need to control for sample selection as well. It is not very difficult to control for sample selection in cross-sections, but it becomes a more complex problem in panel data models with attrition, and the unbalanced nature of the panels begins to be an issue. In this paper we follow the work of Hsiao (1986) and the recommendations in Wooldridge (2002) to account for selection in our unbalanced panel.

Sample selection can be depicted in the following schematic model of retirement expectations that controls for individual clustering (see Deaton 1997):

$$X_{ti}^{e} = \alpha_1 + \gamma_1 \Omega_{t,1i} + \varepsilon_{t,1i}, \qquad (7)$$

and

$$Y_i = \alpha_2 + \gamma_2 \Omega_{2i} + \varepsilon_{2i}, \qquad (8)$$

where the set of individual characteristics, Ω_{1i} consists of various socio-economic and demographic variables, and other variables we will describe below, and X_{ti}^{e} represents the expected retirement age at time *t*. Y_i is an indicator of whether or not any thought has been given to the retirement plans and determines whether or not we observe X_{ti}^{e}, and ε_{1i} and ε_{2i} are not necessarily independent. This means that the selection rule is potentially not independent of the behavioral function being estimated. It is fairly straightforward to estimate the full model by Maximum Likelihood or by standard two-step procedures.

The HRS provides us with repeated observations of the same individuals. This allows us to control for potential unobserved components that could enter our econometric model. In part it reduces the omitted variable

[13] For an interesting discussion of these econometric concerns see Wooldridge (2002), Hsiao (1986), Vella and Verbeek (1999), Vella (1998), and Kyriazidou (1997).

biases present by the fact that the econometrician cannot observe the whole array of variables that affect individual decision making.[14] Then we can write (7) as:

$$X_{ti} = \alpha + \gamma\ \Omega_{ti} + c_i + u_{ti}\,, \qquad (9)$$

where c_i represents the unobserved heterogeneity component, and u_{ti} are the idiosyncratic disturbances. We can estimate this model assuming either no correlation between the observed explanatory variables and the unobserved component (random effects), or allowing for arbitrary correlation between the unobserved effect and the observed explanatory variables (fixed effects). We can then test whether the random effects specification or the fixed effect specification is more appropriate, and whether the former is more valid than the pooled OLS regression.

Then we have to take into account selection and attrition bias in the panel model as well. We combine the specifications in (7) to (9) to include the unobserved component as well as the sample selection correction following Hsiao (1986) and Wooldridge (2002), where we essentially extend Heckman's (1979) strategy to correct for sample selection to a panel data model, but where in the first stage we estimate a pooled-probit of whether individuals have thought about retirement, controlling for individual clustering, and use it as the selection criteria into equation (9).

Another issue we will be concerned with is the potential endogeneity of our regressor of interest, the previous period retirement expectation, in the presence of a classical selection problem.[15] Again here we follow Wooldridge (2002) to consistently estimate the effect of previous expectation on current expectation, and from (4) we can write

$$X^{e}_{t+1,i} = \alpha_1 + \beta X^{e}_{t,i} + \gamma_1 Z_{t,1i} + \varepsilon_{t,1i}\,, \qquad (10)$$

$$X^{e}_{t,i} = \alpha_2 + \gamma_2 Z_{t,2i} + \varepsilon_{t,2i}\,, \qquad (11)$$

and

$$Y_i = \alpha_3 + \gamma_3 Z_{t,3i} + \varepsilon_{3i}\,, \qquad (12)$$

where we first estimate the selection equation (12) using a probit specification, where Y_i is equal to one if both the expectation in period t and the expectation in period $t+1$ are observed. Z_3 in equation (12) includes all the

[14] Given our econometric specifications, we cannot control for the unobserved components that vary over time. However, it is not unreasonable to assume that some of these components would bias the coefficients of interest, γ, upwards while others would bias them downwards, leading to attenuation of the biases.

[15] We are concerned about reporting errors that may be correlated with measurement errors in other factors.

exogenous variables, and any exclusion restriction of the selection equation with respect to the structural equation (10). We then consistently estimate (10) by performing a modified 2SLS procedure, where the first stage includes as instruments all the exogenous variables used in (12), the Inverse Mills' ratio from the probit equation, and any additional instruments, the validity of which will be tested.

4. Data

The HRS is a nationally representative longitudinal survey of 7,700 households headed by an individual aged 51 to 61 as of the first round of interviews in 1992-93. So far five waves of data are available, and we use all of them in our analysis. The primary purpose of the HRS is to study the labor force transitions between work and retirement with particular emphasis on sources of retirement income and health care needs. It is a survey conducted by the Survey Research Center (SRC) at the University of Michigan and funded by the National Institute on Aging. The data for the respondents are merged from wave 5 backwards to waves 4, 3, 2, and 1, and we construct a set of consistent variables on different sources of income, financial and non-financial wealth, health, health insurance, and socio-economic characteristics that will be assigned to each decision maker appropriately.

We include any observation for respondents that are working, full time or part time, in any wave and non-employed (but searching for jobs) that report retirement plans. We exclude respondents who do not report retirement plans for more than two consecutive years and for whom we observe relevant information, which results in more than 11,000 respondents in the sample and around 24,000 person-period observations once missing values in the main variables of interest are considered. We construct relevant dependent and independent variables for each wave. We also construct a transitions dataset that consists of changes to these covariates across waves, since for part of the analysis we are interested in changes of variables over time.[16]

In each wave respondents are asked when they plan to fully or partially depart from the labor force.[17] They are also asked if they thought much about retirement. These questions are not mutually exclusive, but most of the people who have not thought about retirement do not report an expected age.[18] A non-trivial number of individuals report they will never retire, although these same people often change their minds at some point and report an age.

---

[16] If there are missing values for one wave we use the prior wave of information, but we only go back one wave.
[17] In wave 1 they were only asked about a full departure.

The analysis could potentially be sensitive to how we treat "never retire" since we need to put in some older age that we select arbitrarily. We report results for two alternatives. First, we assign an age of 77 for those who never retire (estimated longevity). We call this the full model. Next, we omit them from the analysis, but correct for the selection into this group.[19]

Expected retirement ages are distributed similarly to actual retirement ages with peaks at ages 62 and 65 as well as a peak for the bunching at 77 for those who never plan to retire. Over time these expectations converge to between 62 and 65 with fewer people maintaining plans of retirement before age 62 or after age 65. Table 1a and 1b provide an analysis of how these expectations compare with a number of retirement measures (see also for example Panis 2002). Interestingly, expectations are much more concentrated on the traditional peaks than actual retirement which, except for a measure that uses the age at which individuals start to receive Social Security, is much smoother. In any case, besides the apparent focal points that ages 62 and 65 play when people form expectations, it seems clear that retirement expectations are measuring retirement itself. This is encouraging evidence if we are to use this data to further characterize the process of how these expectations are formed.

As indicators of economic status, we construct variables of net worth and household wealth. We also control for income for the respondent. We use health limitations, self-ratings, as well as a number of disease indicators and activities of daily living to control for health status. We also use the self-reported probability of living to age 85 as a measure of the individual's time horizon, which may be correlated with health status. Hurd and McGarry (1995) find this variable to be highly correlated with own health status and parent mortality.

Table 2 displays descriptive statistics on the pooled sample. It is broken up into three tables. The first two tables show levels by sample type and selection criteria. The next table reports transitions by sample type. For the transition data each observation represents transitions between two of the five waves so that each individual has up to four observations. 73% of our sample has information for all four transition periods and 84% have it for three.

Beginning with Table 2a, we see that removing those who report that they will never retire—around 5% of the full sample—reduces the average expected retirement age from 64.6 to 62.3. This is to be expected since those who were removed are assigned a value of 77. Other than this definitional difference, there are relatively few

[18] Many of them report that they will never retire. If they have not given it any thought, and they say they will never retire, we treat their expected retirement age as missing. If they give a retirement age we treat them as non-missing.
[19] Using age 77 creates a larger variance in the mean of this estimate, which affects the constant in the regressions.

significant differences between the two samples. Earnings of those who plan to never retire are slightly lower, and so is their net wealth, but they are much more likely to be self-employed and not to have health insurance. Also, they are more likely to be male, less likely to be married, and have longer expected time horizons.

Table 2b compares by the selection criteria; whether or not retirement is thought about for the full sample. Roughly 47% of the sample gave retirement some thought. Those who have not thought about retirement are less likely to be employed during the panel, and are financially worse off. They are also more likely to receive government health insurance. Moreover, they report to be in worse overall general health by subjective and objective measures, and are more likely to be female and less educated.

Looking at the transitions of this data in Table 2c, on average people are postponing retirement since the average deviation in expected retirement age is positive. The frequency of change is only slightly higher for the full sample compared with the restricted sample. Plans do change around 55% of the time in the full sample. This is especially true among those that at some point thought they would never retire. This sub-sample is very likely to actually report an expected age in subsequent waves. There are smaller changes in earnings and wealth among those who never plan to retire, but more of them become uninsured. Those, that over time think they will never retire, are faced with much smaller changes in the covariates. This is not unexpected since they are a select sample.

5. Empirical Results

Analysis of Expectations Formation

The first part of the analysis examines factors that influence how people form retirement expectations. As we mentioned above, this is the first paper to study this question using panel data and focusing on the expectation formation with controls for unobserved heterogeneity and sample selection. Table 3a reports pooled OLS, and fixed and random effects results with no sample selection controls for the full sample. Using Hausman's specification test we can reject the random effects model over the fixed effects (we can reject the hypothesis that the coefficients are equal). However, this is not very surprising since many of the variables of interest vary very little creating instability in the parameters of the fixed effect specification. Using a Lagrange Multiplier (Breusch-Pagan) test we find that the random effects specification is preferred to the pooled OLS. We find that people with higher net worth and higher earnings plan to retire earlier probably because they can afford to. Those with private health insurance expect to retire later than those with access to employer provided coverage or retiree coverage, and self-employed individuals plan to retire much later.

It is interesting to note that those with arthritis plan to retire earlier, but most other health factors, surprisingly, are not significant, although this is something that the selection correction will clarify. As individuals grow older they tend to lean toward retiring earlier, as do married people. Males and whites are more likely to expect to retire later.

Controlling for sample selection is key, and changes the results considerably. This is not surprising given the considerable differences in a number of factors whereby people have given thought to retirement or not. We see in table 3b, that after controlling for the selection into thinking about retirement, the effect of earnings is of the opposite sign. So people who earn more are now more likely to plan to work longer (higher opportunity cost, larger substitution effect). Also, the private health insurance effect becomes stronger in the Pooled Corrected OLS but insignificant in the Random Effects model. Interestingly, after we correct for selection, the main health indicators are now significant, with bad health indicators or health limitations making the person more likely to plan to retire earlier, with the exception of having difficulty with climbing stairs, which still lead to plans for later retirement.

After controlling for sample selection, we find that more educated individuals plan to retire later. The effect of age is no longer significant, but the time trends, proxied by wave indicators, show that those reporting expected retirement ages during the bull market of the late 1990s are more likely to expect to retire earlier, a result consistent with the conclusions of Coronado and Perozek (2001). By the sign, size, and high significance of the selection correction term, we can conclude that people with a higher propensity to have thought about retirement, retire later.

Table A.1. in the Appendix shows the determinants of the selection criteria using both cross-section and panel data techniques. We observe a good fit for these two models. Also in the Appendix, Table A.2. reports the selection corrected results for the restricted sample. Removing those who report that they will never retire changes some of the estimated magnitudes. Mainly economic factors (including education) become more important when we remove the ‘nevers’, but the collinear self-employment effect goes away, since ‘nevers’ are more likely to be self-employed. The effect of marital status, which may be an indicator of employer-provided benefits through the spouse, becomes insignificant without the ‘nevers’. Clearly, the individuals that do not plan to retire are an important source of variation to understand how people plan for retirement. While the analysis is sensitive to how we deal with the 'nevers', we believe that they belong in the sample. By excluding them we would remove a select part of the expected retirement age distribution, thereby biasing our results upward. In other words, we would be keeping those more likely to postpone retirement and omitting those more likely to expedite it. Relying on sample selection to deal with observable differences results in a loss of information.[20]

Tests of Rational Expectations

Table 4 reports the weak and strong RE tests for the full sample. The data support the weak and strong RE hypotheses only in an augmented model that corrects for sample selection and measurement error in the report of expected retirement age, resulting in a selection corrected IV specification.[21]

Notice that we perform an F-test based on the null hypothesis that $\beta=1$ in equation (4), to test the RE hypothesis. We obtain coefficients for beta of 1.027 for the weaker test, which given the precision of the estimate cannot reject the hypothesis that expectations follow a random walk. For the pooled OLS estimation the weak test is effectively a unit root test, and as such, and following the literature on testing unit roots in panel data surveyed by Bond, Nauges, and Windmeijer (2002), we have to perform a correction to obtain the appropriate critical value. However, this matters very little since the unit root hypothesis is soundly rejected.

For the strong test we estimate the model of equations (10) to (12), using the Corrected IV procedure. The beta parameter is estimated to be equal to 1.00681, which clearly fails to reject the RE hypothesis. Notice the importance of both instrumenting the previous period's expectations, and controlling for sample selection. We also report in the table tests that show that we cannot reject that we have robust instruments and that the overidentification in the 2SLS is correct.[22] In fact the reported results are the product of robustly estimating the system of equations via GMM, which provides robustness against unknown forms of heterokedasticity.

---

[20] Although, the analysis is somewhat sensitive to the assigned age for 'nevers', sensitivity analysis reveals that the results are not significantly different.

[21] The findings are robust across many specifications and empirical techniques including panel data methods. Much of the individual component is explained by time-invarying variables (there is no remaining individual component in a random effects model if we exclude these covariates). The justification for including these time-invarying components in the test is because they may be correlated with other time-varying unobservables. However, due to collinearity problems it was necessary to remove age from the list of covariates in the second stage of the Corrected IV procedure. As Vella (1998) discusses, this might be due to the fact that age in our sample has a fairly small range, leading to the apparent linearity of the inverse Mills' ratio. We are therefore assuming that age is a proxy for the information set, and only matters in terms of making you more or less likely to think about retirement, but does not directly affect the expected retirement age.

[22] For a discussion of how to test the robustness of instruments and the overidentifying restrictions see Bound, Jaeger, and Baker (1995), Staiger and Stock(1997), Stock, Wright, and Yogo (2002), and Baum, Schaffer and Stillman (2002).

The strong test includes information available at time *t* that should not be significant after controlling for time *t* expectations. Significance would imply that this factor was not incorporated in the previous period's expectations and implies underutilized information. After controlling for sample selection and measurement error we find that most of these factors are no longer significant, in fact only wealth and being male are. The joint hypothesis that all the coefficients are equal to zero cannot be rejected at any traditional level of significance.[23]

The objective behind instrumental variables estimation here is to correct for potential measurement error in the reported expected age of retirement at time *t*. Since people are reporting expectations over uncertain events, we expect some degree of reporting error that may be correlated with unobserved factors. In fact, Bernheim (1988) finds that expectations are reported with noise, and our Table 1a and 1b seem to support that theory as well. Like Bernheim (1990), we correct for this problem using instrumental variables analysis. The instruments must be correlated with the expected retirement age but not with the error term or any new information relevant to the *t+1* expectation. We use time *t* subjective survival to age 85 probabilities and an indicator of smoking behavior as instruments for expected retirement age. In the selection corrected IV, the inverse Mills' ratio is a third instrument, along with the rest of the exogenous variables from the selection equation, as suggested by Wooldridge (2002).[24] The strongest specification is indeed the selection corrected IV on the full sample.[25] Interestingly, this Corrected IV technique seems to circumvent one of the traditional drawbacks of instrumental variables estimation, and that is the large increase in standard errors in the estimates of the instrumented variables. In Table 4 we can observe that when we move from the OLS estimation to the IV estimation standard errors are multiplied by a factor of 8, but once we use the Corrected IV technique they only treble with respect to OLS while the coefficient of interest more than doubles. The importance of the selection correction in this setting, contrasts with the results by Bernheim (1990) where selection was not important, and the inability to reject rationality was in part the product of large standard errors.

---

[23] It is true, however, as in Bernheim (1990) that this is trivially the case if individuals never adjust their expectations. But plenty of adjustment goes on in the data, and it seems implausible that all can be blamed on measurement error.

[24] The exclusion restrictions in the selection equation include indicators for whether the father and mother of the respondent reached retirement age.

[25] Notice that in columns two and three of Table 4, and in Table A.4. we do not report the adjusted R^2 measure of fit. This is common practice, but it is rarely mentioned in empirical work. These types of measures do not have independent significance in structural estimation à la IV, given that we are after estimating population parameters, which we consider invariant to the particular way of identifying the parameters (instruments), not after minimizing a particular prediction problem. See Ruud (2000, p. 515-516 for a discussion)

Table A.3.1. and A.3.2. in the Appendix provide the results for the first stage of the IV specifications for the weak and strong tests respectively, and Table A.4. presents the results for the IV and Corrected IV specifications for the restricted sample. For this sample we cannot reject the RE hypothesis when we control for measurement error, due to the large standard errors but we narrowly can when we use the Corrected IV procedure. This rejection, and the fact that the coefficient of interest is estimated to be larger than 1, is expected since we are removing the upper tail of the distribution of expected retirement ages and are left with a sub-sample that is more likely to postpone their retirement plans.

The Role of New Information

In this part of our analysis we are interested in the role of new information, since the prior period, on expected retirement age. We examine the effects of changes to all of the relevant factors on plans by estimating equation (6) but correcting for sample selection bias and measurement error, following the same strategy as in equations (10) to (12). We hypothesize that there will be no significant effects if, on average, people are able to completely anticipate the new information. Notice, however, that a significant coefficient on any covariate does not indicate surprise, since it is easy to write an example where even if individuals perfectly know the distribution over the uncertain event, once this happens, the estimated coefficient would be positive. A positive coefficient can be interpreted to mean that the individual could have known the probabilities, but not what will happen. Table 5 reports results for the full sample, and in Table A.5. in the Appendix we provide the results for the restricted sample.

The findings are significant and deserve our attention. When we estimate equation (6) and correct for selection bias and measurement error in order to attenuate the biases resulting from the fact that we do not have all the information of the structural model in (5), we find that the coefficient on the expected retirement age reported in the previous period is not significantly different from 1, suggesting that we have greatly reduced the biases, and increased the reliability of the estimated coefficients.[26] The only factors that significantly alter retirement plans are some health conditions and the availability of pensions and private health insurance. Focusing on the selection corrected IV specification, individuals who have been newly diagnosed with diabetes or high blood pressure expect

[26] The error term seems to be more correlated with the expected retirement age as of time *t* than the new information set, since the estimated effects (γ) are not significantly different after correcting for measurement error but β was biased downward without the measurement error correction.

to delay retirement, maybe because they have to slow down and therefore need to work more years to reach the same level of well being, or because their new diagnoses reduces the uncertainty about their future health status and allows them to plan with a longer time horizon. Changes to economic status have no effect, except for those individuals who now have access to a pension through their employer. New pension availability probably proxy for people returning to the labor force, which correlates with expecting to retire later. New private health insurance could proxy for moving into self-employment.

Table A.5. in the Appendix provides the IV selection corrected results for the restricted sample, which presents a very similar picture, with the added results that individuals who have been diagnosed with a stroke since the last wave (last two years) and survive, are expecting to retire earlier by about a year, and those that got married are now expecting to retire earlier, maybe due to access to more resources through the new spouse.

We conclude from these results that researchers need to pay special attention to incorporating the uncertainty of health shocks, and certain economic variables like health insurance, in dynamic models, or models that assume rational expectations. This is something the research community has been relatively aware of for some time, but that requires considerable modeling efforts.

6. Conclusions and Future Research

There are three main areas where this research makes its contributions. First, we examine factors that influence expectation formation using panel data econometric techniques to control for potential sample selection bias and unobserved heterogeneity. We learn that unobserved heterogeneity helps to better explain the variation in expectation formation and that people who have thought about retirement are selectively different in factors that influence retirement. The findings are consistent with prior work on retirement behavior as well as cross sectional work on expectation formation. Health and socio-economic factors are the factors that influence the formation of expectations, with health explaining more of the variation.

Second, we test the Rational Expectations hypothesis in the formation of expectations for retirement. We perform various types of tests and cannot reject the RE hypothesis after controlling for reporting errors and sample selection. This has important implications for a wide variety of models that use this assumption, and for dynamic models that are often identified under rational expectations assumptions.

Finally, we examine the role of new information in retirement expectation formation. This time we focus on how well people anticipate changes over factors relevant to the retirement decision. If we can incorporate uncertainty in a way such that changes to the information set can be modeled following a known distribution the dynamic model is more clearly identified. A model of perfect foresight is rejected with respect to health variables and some economic variables, so on average not all changes are anticipated. We find that some components of health that are associated with various health conditions warrant extra attention in dynamic models or models that assume some type of rationality, if we seek to better fit the data.

The results in this analysis are meant to foster further discussion and research on the issues surrounding the role of expectations in economic modeling. We have provided and tested a methodology, building upon the long RE tradition, which allows researchers to use the ever increasing set of data sources that provide expectations variables.

The econometric work motivates a more structural approach to examining transitions to health states and their effects on plans. We are working on a preliminary dynamic model that incorporates health uncertainties and plan to estimate that model incorporating a health transition model as well as introducing uncertainty to other factors. This is all part of a broader research agenda where we aim to simulate Social Security reform proposals (in the form of changes to benefits) as well as examine the implications of these reforms for savings behavior. Future work will also examine joint expectation formation, and their relationship with joint retirement decisions.[27]

[27] We have done most of the analysis in this paper on married couples, assuming that the spouse's information is exogenous to the individual utility maximization problem. We do find that spouse's information is significant in a model of expectations formation, and therefore it makes sense to expand the model to allow for joint expectation formation. On the other hand, changes to spouse's variables seem to be well anticipated.

References

Baum, C.F., M.E. Schaffer, and S. Stillman (2002): "Instrumental Variables and GMM: Estimation and Testing," Boston College Working Paper 545.

Bernheim, B.D. (1988): "Social Security benefits: An empirical study of expectations and realizations." In *Issues in Contemporary Retirement.* Ed. E. Lazear and R. Ricardo-Campbell. 312—345. Palo Alto: Hoover Institution.

Bernheim, B.D. (1989): "The timing of retirement: A comparison of expectations and realizations." In *The Economics of Aging*. Ed. D. Wise. Chicago: University of Chicago Press, 335—355.

Bernheim, B.D. (1990): "How Do the Elderly Form Expectations? An Analysis of Responses to New Information." In *Issues in the Economics of Aging*. Ed. D. Wise. Chicago: University of Chicago Press, 259—285.

Bond, S., C. Nauges, and F. Windmeijer (2002): "Unit Root and Identification in Autoregressive Panel Data Models: A Comparison of Alternative Tests." IFS Working Paper.

Bound, J., D.A. Jaeger, and R.M. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, **90-430** 443—450.

Coronado, J.L., and M. Perozek (2001): "Wealth Effects and the Consumption of Leisure: Retirement Decisions During the Stock Market Boom of the 1990s," manuscript. Federal Reserve Board.

Christiansen, C. (2003): "Testing the expectations hypothesis using long-maturity forward rates," *Economics Letters*, 78 175—180.

Davies A., and K. Lahiri (1999): "Re-examining the rational expectations hypothesis using panel data on multi-period forecasts," *in Analysis of panels and limited dependent variable models*. Edited by Cheng Hsiao, Kajal Lahiri, Lung-Fei Lee, and M. Hashem Pesaran. Cambridge University Press.

Deaton, A. (1997): *The Analysis of Household Surveys*. Baltimore: Johns Hopkins Press

de Leeuw, F., and M.J. McKelvey (1981): "Price Expectations of Business Firms," *Brookings Papers on Economic Activity*, **1981-1** 299—314.

de Leeuw, F., and M.J. McKelvey (1984): "Price Expectations of Business Firms: Bias in the Short and Long Run," *American Economic Review*, **14** 99—110.

Dominitz, J., and C.F. Manski (1996): "Eliciting Student Expectations of the Returns to Schooling," *Journal of Human Resources*, **31-1** 1—26.

Dominitz, J., and C.F. Manski (1997): "Using expectations data to study subjective income expectations," *Journal of the American Statistical Association*, **92** 855—867.

Dominitz, J., C.F. Manski, and J. Heinz (2001): "Social Security Expectations and Retirement Savings Decisions," manuscript, Carnegie Mellon University and Northwestern University.

Dwyer, D. S. (2002): "Planning for Retirement: The Role of Health Shocks," manuscript. SUNY-Stony Brook.

Dwyer, D.S., and O.S. Mitchell (1999): "Health Shocks as Determinants of Retirement: Are Self-Rated Measures Endogenous?" *Journal of Health Economics*, **18** 173—193.

Dwyer, D.S., and J. Hu. (1999): "The Relationship Between Retirement Expectations and Realizations: The Role of Health Shocks," in *Forecasting Retirement Needs and Retirement Wealth*. Olivia Mitchell, P. Brett Hammond, and Anna M. Rappaport (eds.). University of Pennsylvania Press.

Evans, G.W., and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press: Princeton and Oxford.

Fair, R.C. (1993): "Testing the Rational Expectations Hypothesis in Macroeconomic Models," *Oxford Economic Papers*, **45-2** 169—190.

Figlewski, S., and P. Wachtel (1981): "The Formation of Inflationary Expectations," *Review of Economics and Statistics,* **63-1** 1—10.

Figlewski, S., and P. Wachtel (1983): "Rational Expectations, Information Efficiency, and Tests using Survey Data," *Review of Economics and Statistics,* **65-3** 529—531.

Forni, L. (2002): "Expectations and Outcomes," manuscript. Central Bank of Italy.

Gramlich, E.M. (1983): "Models of Inflation Expectations Formation: A Comparison of Household and Economic Forecasts," *Journal of Money, Credit, and Banking*, **15** 155—173.

Heckman, J.J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, **47-1** 153—161

Hsiao, C. (1986): *Analysis of Panel Data*. Econometric Society Monographs. Cambridge University Press.

Hurd, M., and K. McGarry (1995): "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *Journal of Human Resources*, **30** S268—92.

Hurd, M., and M. Retti (2001): "The Effects of Large Capital Gains on Work and Consumption: Evidence from four waves of the HRS," manuscript. RAND Corporation. Santa Monica.

Kimball Dietrich, J., and D.H. Joines (1983): "Rational Expectations, Information Efficiency, and Tests using Survey Data: A comment," *Review of Economics and Statistics,* **65-3** 525—529

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, **65-6** 1335—1364.

Leonard, J.S. (1982): "Wage Expectations in the Labor Market: Survey Evidence on Rationality," *Review of Economics and Statistics*, **64** 157—161.

Lovell, M.C. (1986): "Tests of Rational Expectations Hypothesis," *American Economic Review*, **76-1** 110—124.

Lucas, R.E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, **4** 103—124.

Lumsdaine, R.L., and O.S. Mitchell (1999): "New Developments in the Economic Analysis of Retirement," *Handbook of Labor Economics*, **3C** 3261—3307.

Lusardi, A (1999): "Information, Expectations, and Savings for Retirement", in *Behavioral Dimensions of Retirement Economics*, by Henry J Aaron Brookings Institute.

McCallum, B.T. (1980): "Rational Expectations", edited volume of the *Journal of Money, Credit and Banking*. **12-4.**

Muth, J.F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, **29-3** 315—335.

Panis, C. (2002): “Retirement Planning,” manuscript. RAND Corporation: Santa Monica.

Pesaran, M. Hashem. (1987): *The Limits to Rational Expectations*. Basil Blackwell: Oxford and New York.

Prescott, E. (1977): “Should Control Theory be Used for Economic Stabilization? *Journal of Monetary Economics*, **Suppl**. 13—38.

Rosen, S. (1990): “Comment on How Do Elderly Form Expectations? An Analysis of Responses to New Information.” In *Issues in the Economics of Aging*. Ed. David Wise. Chicago: University of Chicago Press.

Ruud, (2000): *An Introduction to Classical Econometric Theory*. Oxford University Press: New York and Oxford.

Sargent, T.J., and N. Wallace (1976): “Rational Expectations and the Theory of Economic Policy,” *Journal of Monetary Economics*, **2** 169—183.

Sargent, T.J. (1995): “Adaptation of Macro Theory to Rational Expectations,” manuscript. Hoover Institution and University of Chicago.

Schuetz, A. (1943): “The Problem of Rationality in the Social World,” *Economica*, New Series **10-38** 130—149.

Simon, H.A. (1979): “Rational Decision Making in Business Organizations,” *American Economic Review*, **69-4** 493—513.

Staiger, D., and J.H.Stock (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, **65-3** 557—586.

Stock, J.H., J.H. Wright, and M. Yogo (2002): “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economics Statistics*, **20-4** 518—529.

Tobin, J. (1980): *Asset Accumulation and Economic Activity*. The University of Chicago Press.

van der Klaauw (2000): “On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices,” manuscript, UNC-Chapel Hill.

Vella, F. (1998): “Estimating Models with Sample Selection Bias: A Survey,” *Journal of Human Resources*, **33-1** 127—169.

Vella, F., and M. Verbeek (1999): “Two-step estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias,” *Journal of Econometrics*, **90** 239—263.

Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Young, W., and W. Darity (2001): “The Early History of Rational and Implicit Expectations,” *History of Political Economy*, **33-4** 773—813.

Table 1.a. Distribution of Retirement Expectations and Actual Retirement. Full Sample

	W1	W2	W3	W4	W5	Self-R Ret.	Social Sec.	Fully Retired	Full and Partial Retirement
Age <50	0.19%	0.25%	0.14%	-	0.23%	9.11%	0.7%	11.8%	13.1%
Age 50	0.54%	0.34%	0.39%	0.22%	0.11%	2.1%	0.2%	2.51%	2.73%
Age 51-54	1.76%	1.14%	0.97%	0.9%	0.69%	11.3%	1.13%	12.7%	14.1%
Age 55	6.07%	4.7%	2.64%	2.24%	1.25%	6.4%	1%	6.18%	6.54%
Age 56-59	7.09%	7.34%	7.19%	5.43%	3.37%	**20.5%**	5.34%	**21.2%**	**21.5%**
Age 60	8.71%	9.71%	7.73%	7.13%	4.68%	7.4%	4.91%	7.22%	6.81%
Age 61	1.24%	1.1%	1.53%	1.48%	1.31%	7.51%	5.57%	6.56%	6.37%
Age 62	**30.8%**	**29.9%**	**27.5%**	**27.3%**	**22.6%**	**15.8%**	**49.4%**	**13.5%**	**12.9%**
Age 63-64	3.62%	4.79%	4.66%	5.29%	5.99%	9.63%	16.4%	9.18%	7.84%
Age 65	**19.2%**	**20.4%**	**23.1%**	**23%**	**23.3%**	**5.67%**	**10.6%**	**4.77%**	**4.54%**
Age >65	20.8%	20.3%	24.3%	27%	35.4%	4.54%	4.63%	4.36%	3.45%
Never	17.1%	15.8%	17.1%	14.8%	17.3%	-	-	-	-
# Obs.	3,708	3,256	2,808	2,230	1,753	5,346	3,967	4,902	6,404

W1 to W5: the expected retirement age reported in the respective rounds of data.
Self-R Ret.: reported retirement age when individuals are asked about their employment status.
Social Sec.: age at which they started to receive Social Security benefits.
Fully Retired: answer to a direct question regarding when did they fully withdrawn from the labor force.
Full and Partial Retirement: includes partial retirement in the above definition.

Table 1.b. Distribution of Retirement Expectations and Actual Retirement. Restricted Sample

	W1	W2	W3	W4	W5	Self-R Ret.	Social Sec.	Fully Retired	Full and Partial Retirement
Age <50	0.26%	0.29%	0.17%	-	0.28%	9.11%	0.7%	11.8%	13.1%
Age 50	0.65%	0.4%	0.47%	0.26%	0.14%	2.1%	0.2%	2.51%	2.73%
Age 51-54	2.12%	1.35%	1.16%	1.06%	0.83%	11.3%	1.13%	12.7%	14.1%
Age 55	7.32%	5.58%	3.18%	2.63%	1.52%	6.4%	1%	6.18%	6.54%
Age 56-59	8.56%	8.72%	8.68%	6.36%	4.07%	**20.5%**	5.34%	**21.2%**	**21.5%**
Age 60	10.5%	11.5%	9.32%	8.36%	5.66%	7.4%	4.91%	7.22%	6.81%
Age 61	1.5%	1.31%	1.85%	1.74%	1.59%	7.51%	5.57%	6.56%	6.37%
Age 62	**37.1%**	**35.6%**	**33.1%**	**32.1%**	**27.3%**	**15.8%**	**49.4%**	**13.5%**	**12.9%**
Age 63-64	4.36%	5.7%	5.63%	6.21%	8.49%	9.63%	16.4%	9.18%	7.84%
Age 65	**23.2%**	**24.2%**	**27.8%**	**27%**	**28.2%**	**5.67%**	**10.6%**	**4.77%**	**4.54%**
Age >65	4.42%	5.36%	8.68%	14.4%	21.9%	4.54%	4.63%	4.36%	3.45%
# Obs.	3,374	2.740	2,328	1.901	1,449	5,346	3,967	4,902	6,404

W1 to W5: the expected retirement age reported in the respective rounds of data.
Self-R Ret.: reported retirement age when individuals are asked about their employment status.
Social Sec.: age at which they started to receive Social Security benefits.
Fully Retired: answer to a direct question regarding when did they fully withdrawn from the labor force.
Full and Partial Retirement: includes partial retirement in the above definition.

Table 2a. Summary Statistics. Variables in levels

Variables	Full Sample N=23,669	Restricted Sample N=22,000	Never Retire N=1,669
Retirement Plans and Outcomes			
Expected retirement age	64.584(6.478)	62.378(4.143)	77(0.00)
Employee	0.794(0.405)	0.802(0.399)	0.689(0.463)
Self employed	0.173(0.378)	0.163(0.369)	0.311(0.463)
Financially Knowledgeable	0.657(0.475)	0.656(0.475)	0.678(0.467)
Economic factors			
Net worth (in $100,000)	2.449(5.181)	2.434(5.123)	2.634(5.892)
Housing wealth (in $100,000)	0.769(1.248)	0.771(1.267)	0.742(.966)
Respondent' Income (in $1,000)	29.213(54.304)	29.325(54.650)	27.74(49.51)
Has a private pension	0.593(0.491)	0.605(0.489)	0.445(0.4971)
Health Insurance			
Employer provided	0.699(0.459)	0.703(0.457)	0.635(0.481)
Retiree	0.814(0.389)	0.812(0.391)	0.842(0.364)
Government	0.082(0.274)	0.078(0.267)	0.134(0.341)
Private	0.188(0.391)	0.186(0.389)	0.217(0.412)
No health insurance	0.087(0.282)	0.083(0.276)	0.132(0.338)
Health factors			
Health limitation	0.187(0.390)	0.187(0.390)	0.183(0.386)
Good-Very Good-Excellent Health	0.866(0.340)	0.867(0.340)	0.864(0.343)
Doctor visits	5.191(7.075)	5.210(7.065)	4.946(7.212)
Probability of living to age 85	0.470(0.306)	0.466(0.305)	0.521(0.321)
High blood pressure	0.228(0.419)	0.228(0.420)	0.219(0.414)
Diabetes	0.060(0.238)	0.060(0.238)	0.061(0.239)
Arthritis	0.283(0.450)	0.283(0.450)	0.278(0.448)
Difficulty walking multiple blocks	0.082(0.275)	0.082(0.274)	0.083(0.276)
Difficulty climbing stairs	0.047(0.212)	0.046(0.210)	0.058(0.234)
Stroke	0.003(0.052)	0.003(0.053)	0.001(0.034)
Heart problems	0.075(0.263)	0.075(0.263)	0.078(0.268)
Cancer	0.007(0.083)	0.007(0.083)	0.007(0.084)
Smoke	0.219(0.414)	0.216(0.411)	0.264(0.441)
Demographic factors			
Age	57.197(5.222)	57.104(5.188)	58.41(5.51)
Male	0.465(0.499)	0.458(0.498)	0.546(0.498)
Married	0.794(0.405)	0.795(0.404)	0.773(0.418)
Bachelor's degree	0.270(0.444)	0.269(0.444)	0.273(0.445)
Professional degree	0.101(0.301)	0.101(0.302)	0.097(0.297)
Mother reached retirement age	0.714(0.452)	0.712(0.453)	0.74(0.438)
Father reached retirement age	0.596(0.491)	0.594(0.491)	0.622(0.485)

Table 2b. Summary Statistics by Sample Selection. Full Sample in levels

Variables	Thought About N=11,062	Not Thought N= 12,607
Retirement Plans and Outcomes		
Expected retirement age	64.584(6.478)	-
Employee	0.840(0.367)	0.753(0.431)
Self employed	0.160(0.367)	0.185(0.388)
Financially Knowledgeable	0.670(0.470)	0.647(0.478)
Economic factors		
Net worth (in $100,000)	2.612(5.484)	2.305(4.895)
Housing wealth (in $100,000)	0.798(1.278)	0.744(1.221)
Respondent's Income (in $1,000)	33.583(67.258)	25.379(39.192)
Has a private pension	0.657(0.475)	0.538(0.499)
Health Insurance		
Employer Provided	0.748(0.434)	0.652(0.476)
Retiree	0.804(0.397)	0.823(0.382)
Government	0.065(0.247)	0.095(0.294)
Private	0.184(0.388)	0.191(0.393)
No health insurance	0.063(0.244)	0.107(0.309)
Health factors		
Health limitation	0.185(0.388)	0.189(0.391)
Good-Very Good-Excellent Health	0.872(0.334)	0.861(0.346)
Doctor visits	5.311(7.054)	5.086(7.093)
Probability of living to age 85	0.468(0.305)	0.472(0.307)
High blood pressure	0.233(0.423)	0.223(0.416)
Diabetes	0.061(0.239)	0.060(0.238)
Arthritis	0.285(0.452)	0.280(0.449)
Difficulty walking multiple blocks	0.081(0.272)	0.083(0.277)
Difficulty climbing stairs	0.048(0.213)	0.047(0.211)
Stroke	0.003(0.051)	0.003(0.053)
Heart Problems	0.077(0.266)	0.074(0.261)
Cancer	0.008(0.087)	0.006(0.079)
Smoke	0.202(0.402)	0.234(0.423)
Demographic factors		
Age	57.558(4.824)	56.880(5.528)
Male	0.506(0.500)	0.428(0.495)
Married	0.802(0.398)	0.786(0.410)
Bachelor's degree	0.290(0.454)	0.252(0.434)
Professional degree	0.119(0.323)	0.085(0.280)
Mother reached retirement age	0.728(0.445)	0.702(0.457)
Father reached retirement age	0.603(0.489)	0.590(0.492)

Table 2c. Summary Statistics. Variables in changes

Variables	Full Sample N= 14,028	Restricted Sample N= 12,345	Never Retire N=1,683
Retirement Plans and Outcomes			
Changes in Expected Retirement Age	0.291(5.764)	0.388(3.054)	-0.212(12.550)
Frequency of change	0.544(0.498)	0.508(0.500)	0.731(0.444)
Economic factors			
Changes in Net Worth (in $100,000)	0.061(3.877)	0.066(3.870)	0.024(3.930)
Changes in Income (in $1,000)	0.867(56.914)	0.957(58.874)	0.207(39.683)
Private pension transitions	0.020(0.286)	0.020(0.285)	0.021(0.293)
Health Insurance			
Changes in Private Health Insurance	0.010(0.440)	0.011(0.439)	-0.001(0.448)
Changes in No Health Insurance	-0.008(0.245)	-0.006(0.241)	-0.020(0.276)
Health factors			
Health limitation transitions[28]	0.005(0.402)	0.005(0.397)	0.005(0.435)
Good-V. Gd.-Ex. Health transitions[29]	-0.022(0.362)	-0.020(0.364)	-0.033(0.344)
Stroke transitions[30]	0.002(0.055)	0.001(0.055)	0.003(0.054)
Heart Problems transitions	0.020(0.244)	0.019(0.240)	0.025(0.273)
Cancer transitions	-0.003(0.110)	-0.003(0.109)	-0.004(0.114)
Diabetes transitions	0.005(0.227)	0.004(0.229)	0.009(0.214)
High blood pressure transitions	0.014(0.454)	0.014(0.454)	0.018(0.456)
Arthritis transitions	0.008(0.524)	0.007(0.521)	0.015(0.544)
Smoking transitions	-0.017(0.223)	-0.016(0.218)	-0.024(0.256)
Probability of living to age 85	0.010(0.290)	0.011(0.289)	0.004(0.297)
Doctor visits	1.128(7.700)	1.111(7.717)	1.256(7.573)
Difficulty climbing stairs	0.012(0.230)	0.012(0.230)	0.015(0.235)
Difficulty walking multiple blocks	0.017(0.284)	0.018(0.282)	0.007(0.300)
Demographic factors			
Marriage Transitions[31]	0.006(0.080)	0.006(0.077)	0.010(0.100)

[28] =1 if new limitation, 0 if no change, and -1 if you got better
[29] negative means health worsened
[30] positive means condition (Stroke, cancer, high blood pressure) worsened (0 no change, -1 if better)
[31] =1 means new marriage, = 0 no change, =-1 dissolved marriage (widow, divorce…)

Table 3a. Expectations Formation. Full Sample. No Selection Correction
Coefficient (Standard Error)

Variables	Pooled OLS	Fixed Effects	Random Effects
Economic factors			
Net Worth (in $100,000)	-0.055(0.013)**	-0.001(0.019)	-0.043(0.012)**
Respondent Income (in $1,000)	-0.002(0.001)**	-0.002(0.001)*	-0.003(0.001)**
No Health Insurance	1.866(0.290)**	0.055(0.364)	1.554(0.231)**
Private Health Insurance	-0.025(0.156)	0.356(0.177)**	0.110(0.137)
Self-employed	1.784(0.230)**	1.417(0.386)**	1.876(0.188)**
Pension	-1.358(0.155)**	0.934(0.320)**	-1.266(0.148)**
Financially Knowledgeable	0.213(0.146)	-0.081(0.556)	0.188(0.153)
Health factors			
Health limitation	-0.109(0.162)	-0.477(0.206)**	-0.238(0.151)
Good-V.Good-Exc. Health	0.127(0.173)	0.455(0.226)**	0.195(0.165)
Doctor visits	-0.002(0.009)	-0.009(0.010)	-0.006(0.008)
High blood pressure	-0.128(0.140)	0.141(0.206)	-0.056(0.141)
Diabetes	0.009(0.235)	0.982(0.368)**	0.247(0.237)
Cancer	-0.164(0.619)	0.558(0.793)	0.085(0.593)
Stroke	-0.697(0.773)	-0.788(1.279)	-0.710(0.966)
Heart problems	-0.202(0.212)	0.071(0.307)	-0.159(0.214)
Arthritis	-0.386(0.134)	-0.050(0.183)	-0.294(0.130)**
Difficulty walking multiple blocks	-0.172(0.228)	-0.207(0.296)	-0.155(0.215)
Difficulty climbing stairs	0.331(0.293)	-0.093(0.361)	0.130(0.262)
Demographic factors			
Age	-1.666(0.227)**	-	-1.826(0.172)**
Age Sq.	0.019(0.002)**	-	0.020(0.002)**
Male	0.582(0.137)**	-	0.519(0.151)**
White	0.236(0.156)**	-	0.166(0.167)
Bachelor's Degree	0.385(0.174)**	-	0.374(0.190)**
Professional Degree	-0.340(0.245)	-	-0.330(0.271)
Married	-0.860(0.175)**	-0.009(0.459)	-0.822(0.180)**
Wave 1	0.372(0.165)**	-1.958(0.1701)**	0.422(0.156)**
Wave 2	0.0048(0.165)	-1.523(0.1852)**	0.1982(0.156)
Wave 3	0.1077(0.1459)	-0.794(0.151)**	0.2717(0.1402)*
Constant	99.403(6.708)**	64.78(0.6674)**	104.872(4.949)**
Adj. R^2	0.1871	0.0089	0.1861
Fraction of variance due to unobserved component	-	0.65	0.45
Corr(c_i, ε_i)	-	-0.072	0 (assumed)
Number of Observations	12,401	12,401	12,401

Breusch-Pagan Lagrange Multiplier (OLS vs. RE): $\chi^2(1)=593.11$, P-value=0.0000
Hausman Test Statistic (RE vs. FE): $\chi^2(22)=118.92$, P-value=0.0000.

Table 3b. Expectations Formation. Selection Correction

Variables	Corrected Pooled OLS	Corrected Random Effects
Economic factors		
Net Worth (in $100,000)	-0.044(0.015)**	-0.022(0.014)
Respondent Income (in $1,000)	0.003(0.001)**	0.001(0.001)
No Health Insurance	0.277(0.567)	0.199(0.458)
Private Health Insurance	-0.232(0.173)**	-0.074(0.147)
Self-employed	1.642(0.246)**	1.750(0.199)**
Pension	0.505(0.518)	0.377(0.448)
Financially Knowledgeable	0.240(0.154)	0.247(0.161)
Health factors		
Health limitation	-0.215(0.166)	-0.348(0.160)**
Good-V.Good-Exc. Health	0.352(0.196)*	0.371(0.183)**
Doctor visits	-0.003(0.009)	-0.005(0.008)
High blood pressure	0.034(0.149)	0.113(0.149)
Diabetes	-0.124(0.246)	0.237(0.255)
Cancer	0.368(0.652)	0.497(0.650)
Stroke	-2.269(0.768)**	-2.335(1.101)**
Heart problems	-0.310(0.221)	-0.341(0.228)
Arthritis	-0.020(0.167)	-0.089(0.145)
Difficulty walking multiple blocks	-0.176(0.242)	-0.228(0.229)
Difficulty climbing stairs	1.132(0.369)**	1.053(0.357)**
Demographic factors		
Age	0.284(0.585)	0.054(0.530)
Age Sq.	0.002(0.005)	0.004(0.004)
Male	1.437(0.302)**	1.234(0.273)**
White	0.063(0.181)	0.098(0.184)
Bachelor's Degree	0.364(0.181)**	0.364(0.197)*
Professional Degree	0.658(0.354)*	0.576(0.358)
Married	-0.365(0.217)*	-0.377(0.214)*
Wave 1	0.3532(0.1753)**	0.4184(0.1652)**
Wave 2	0.7724(0.255)**	0.8458(0.2275)**
Wave 3	0.7402(0.222)**	0.8252(0.1997)**
Constant	29.183(20.44)	38.771(18.174)**
Inverse Mills' ratio	11.330(3.156)**	8.813(2.352)**
Adj. R^2	0.1802	0.1784
Fraction due to unobserved component	-	0.46
Number of Observations	11,148	11,148

Table 4. Tests of Rational Expectations- Full Sample

Variables	Pooled OLS	IV	Corrected IV
Weak RE Test (H_0: Beta=1):	***Reject***	***Reject***	***Cannot Reject***
Constant	31.112(1.170)**	20.388(11.655)*	-1.368(2.479)
Expected Retirement Age_t	0.520(0.018)**	0.687(0.183)**	1.027(0.039)**
Test of Over-Id Restrictions	-	Cannot Rej. P-v=.7710	Cannot Rej. P-v=.2347
Test of Weak Instruments	-	Reject P-v=.0000	Reject P-v=.0000
Strong RE Test (H_0: Beta =1):	***Reject***	***Reject***	***Cannot Reject***
Constant	20.725(1.347)**	10.345(6.655)	0.304(5.150)
Expected Retirement Age_t	0.390(0.021)**	0.673(0.170)	1.00681(0.07994)**
Economic factors at time t			
Net Worth (in $100,000)	0.003(0.016)	0.019(0.019)	0.044(0.020)**
Respondent Income (in $1,000)	-0.001(0.001)	-0.001(0.001)	-0.001(0.001)
No Health Insurance	0.918(0.396)**	0.505(0.584)	-0.234(0.540)
Private Health Insurance	0.014(0.191)	-0.066(0.207)	-0.208(0.233)
Self-employed	0.869(0.277)**	0.434(0.323)	0.038(0.309)
Pension	-0.821(0.182)**	-0.488(0.254)	-0.152(0.232)
Financially Knowledgeable	0.012(0.169)	-0.066(0.171)	-0.149(0.188)
Health factors at time t			
Health limitation	0.108(0.200)	0.030(0.210)	-0.061(0.242)
Good-V.Good-Exc. Health	-0.364(0.253)	-0.231(0.255)	-0.274(0.287)
Doctor visits	-0.004(0.010)	0.001(0.010)	0.006(0.012)
High blood pressure	-0.115(0.178)	-0.106(0.189)	-0.095(0.223)
Diabetes problems	-0.497(0.291)*	-0.518(0.318)	-0.445(0.381)
Cancer	-1.552(0.533)**	-1.108(0.652)*	-0.598(0.780)
Stroke	-0.944(0.612)	-0.115(0.726)	1.060(0.722)
Heart Problems	0.002(0.291)	0.049(0.333)	0.063(0.393)
Arthritis	0.000(0.171)	0.021(0.184)	0.119(0.204)
Difficulty walking multiple blocks	-0.477(0.287)*	-0.370(0.316)	-0.226(0.369)
Difficulty climbing stairs	0.282(0.375)	0.356(0.383)	0.328(0.444)
Demographic factors at time t			
Age	0.340(0.022)**	0.200(0.078)**	-
White	0.013(0.182)	-0.093(0.198)	-0.201(0.215)
Male	0.526(0.161)**	0.511(0.169)**	0.462(0.202)**
Bachelor’s Degree	0.425(0.195)**	0.376(0.189)**	0.284(0.209)
Professional Degree	-0.560(0.241)**	-0.466(0.224)**	-0.336(0.245)
Married	-0.306(0.192)	-0.159(0.221)	0.021(0.231)
Wave 1-2	0.1625(0.1701)	-0.0016(0.196)	-0.313(0.212)
Wave 2-3	0.1912(0.1866)	0.182(0.200)	0.135(0.237)
Adj. R^2	0.328	-	-
Test of joint Significance of Covariates	Reject. P-v=.000	Reject P-v=.008	Cannot Rej. P-v=.2645
Test of Over-Id Restrictions	-	Cannot Rej. P-v=.5773	Cannot Rej. P-v=.1257
Test of Weak Instruments	-	Reject P-v=.0000	Reject P-v=.0000
Number of Observations	4,987	4,721	4,631

Table 5. New information and plans for retirement. Full Sample

Variables	Pooled OLS	IV	Corrected IV
Expected Retirement Age$_t$	0.506(0.020)**	0.9223(0.06197)**	0.95988(0.04381)**
Constant	31.931(1.273)**	5.298(3.93)	2.9086(2.7823)
Economic Factors			
Changes in Net Worth (in $100K)	-0.009(0.018)	-0.000(0.019)	-0.000(0.019)
Changes in Earnings (in $1,000)	-0.000(0.001)	-0.000(0.001)	0.000(0.001)
Changes in Priv. Health Insurance	0.192(0.183)	0.441(0.209)**	0.485(0.211)**
Changes in No Insurance	-0.571(0.458)	-0.797(0.522)	-0.776(0.545)
Changes in Pension	0.576(0.271)**	0.684(0.334)**	0.632(0.346)*
Health Factors			
Health limitation transitions[32]	-0.250(0.196)	-0.287(0.220)	-0.313(0.226)
Good-V.Gd-Exc.Hlt. transitions[33]	0.346(0.227)	0.101(0.259)	0.095(0.267)
Stroke transitions[34]	-0.585(0.906)	-1.115(0.926)	-1.375(0.942)
Heart problems transitions	0.144(0.300)	0.149(0.344)	0.095(0.360)
Cancer transitions	1.240(0.653)*	0.739(0.689)	0.762(0.717)
Diabetes transitions	1.057(0.365)**	0.922(0.430)**	0.930(0.455)**
High blood pressure transitions	0.243(0.174)	0.395(0.203)*	0.417(0.209)**
Arthritis transitions	-0.115(0.153)	-0.081(0.178)	-0.075(0.182)
Doctor visits	-0.003(0.009)	-0.013(0.010)	-0.016(0.011)
Difficulty climbing stairs	0.358(0.323)	0.064(0.388)	0.075(0.395)
Difficulty walking multiple blocks	0.069(0.323)	-0.406(0.366)	-0.407(0.377)
Demographic Factors			
Marriage transitions[35]	-0.359(1.030)	-0.609(1.070)	-1.256(1.040)
Inverse Mills' Ratio First Stage	-	-	-27.293(2.371)**
Adj. R^2	0.2512	-	-
Test of Over-Id Restrictions	-	Cannot Rej. P-v=.2029	Cannot Rej. P-v=.3302
Test of Weak Instruments	-	Reject P-v=.0000	Reject P-v=.0000
Number of Observations	4,555	4,257	4,174

[32] =1 if new limitation, 0 if no change, and -1 if you got better
[33] positive means health improved
[34] positive means condition (Stroke, cancer, high blood pressure) worsened (0 no change, -1 if better)
[35] =1 means new marriage, = 0 no change, =-1 dissolved marriage (widow, divorce…)

Table A.1. Selection Equation Results - Probability of Thinking about Retirement

Variables	Probit	Marg. Effects	RE Probit	Marginal Effects
Economic Factors				
Net wealth (in $100,000)	0.001(0.002)	0.001	0.003(0.002)	0.001
Income (in $1,000)	0.002(0.000)**	0.001	0.002(0.000)**	0.001
No Health Insurance	-0.21(0.034)**	-0.082	-0.229(0.039)**	-0.089
Private Health Insurance	-0.028(0.023)	-0.011	-0.026(0.026)	-.010
Self-Employed	-0.024(0.028)	-0.010	-0.026(0.032)	-0.010
Pension	0.233(0.022)**	0.092	0.262(0.025)**	0.103
Financially Knowledgeable	-0.004(0.023)	-0.002	0.000(0.026)	0.000
Health Factors				
Health limitation	-0.002(0.025)	-0.001	-0.014(0.028)	-0.006
Good-V.Good-Exc. Health	0.033(0.028)	0.013	0.024(0.031)	0.009
Doctor visits	0.001(0.001)	0.000	0.001(0.001)	0.000
Probability of living to 85	-0.004(0.030)	-0.001	0.020(0.034)	0.008
Diff. walking multiple blocks	-0.005(0.036)	-0.002	-0.005(0.041)	-0.002
Diff. climbing stairs	0.104(0.044)	0.041	0.148(0.050)**	0.059
High blood pressure	0.010(0.023)	0.004	0.015(0.026)	0.006
Diabetes	0.006(0.039)	0.002	0.017(0.044)	0.007
Cancer	0.131(0.102)	0.052	0.149(0.113)	0.059
Stroke	-0.128(0.156)	-0.050	-0.192(0.185)	-0.075
Heart problems	-0.007(0.036)	-0.003	-0.014(0.040)	-0.005
Arthritis	0.044(0.022)	0.018	0.033(0.024)	0.013
Demographic Factors				
Age	0.248(0.026)**	0.099	0.296(0.029)**	0.117
Age squared	-0.002(0.000)**	-0.001	-0.002(0.000)**	-0.001
Male	0.121(0.023)**	0.048	0.134(0.026)**	0.053
White	-0.040(0.025)	-0.016	-0.034(0.029)	-0.014
Bachelor's degree	-0.007(0.028)	-0.003	-0.003(0.032)	-0.001
Professional degree	0.121(0.041)**	0.048	0.143(0.046)**	0.057
Married	0.056(0.027)**	0.022	0.065(0.030)**	0.026
Mother reached retirement age	0.016(0.022)	0.006	0.022(0.026)	0.009
Father reached retirement age	0.001(0.020)	0.000	0.004(0.023)	0.002
Wave 1	-0.012(0.025)	-0.005	-0.0184(0.0282)	-0.0073
Wave 2	0.1007(0.0026)**	0.040	0.108(0.0295)**	0.0432
Wave 3	0.0855(0.0238)**	0.0340	0.0937(0.027)**	0.0372
Constant	-7.796(0.7489)**	-	-9.315(0.822)**	-
Predicted Probability	0.46611		0.4586	
Log Likelihood	-16048.77		-15689.41	
Pseudo-R^2	0.0266		0.0215	
Number of Observations	23,860		23,860	

Table A.2. Expectations Formation. Restricted Sample

Variables	Corrected Pooled OLS	Corrected Random Effects
Economic Factors		
Net Worth (in $100,000)	-0.048(0.009)**	-0.038(0.009)**
Respondent Income (in $1,000)	0.002(0.001)**	0.001(0.001)*
No Health Insurance	-0.283(0.370)	-0.395(0.317)
Private Health Insurance	-0.212(0.108)**	-0.088(0.092)
Self-employed	-0.026(0.215)	0.090(0.195)
Pension	0.409(0.304)	0.581(0.276)**
Financially Knowledgeable	0.295(0.100)**	0.285(0.102)**
Health Factors		
Health limitation	0.086(0.105)	0.018(0.099)
Good-V.Good-Exc. Health	0.211(0.111)*	0.253(0.114)**
Doctor visits	-0.008(0.005)	-0.006(0.005)
High blood pressure	0.158(0.101)	0.138(0.094)
Diabetes	0.114(0.166)	0.161(0.159)
Cancer	-0.008(0.312)	-0.081(0.399)
Stroke	-0.512(0.550)	-0.967(0.634)
Heart Problems	-0.334(0.139)**	-0.315(0.142)**
Arthritis	-0.187(0.097)*	-0.143(0.088)
Difficulty walking multiple blocks	-0.089(0.150)	-0.093(0.144)
Difficulty climbing stairs	0.056(0.189)	0.168(0.190)
Demographic Factors		
Age	-0.468(0.369)	-0.638(0.331)*
Age Sq.	0.008(0.003)**	0.010(0.003)**
Male	0.625(0.130)**	0.637(0.128)**
White	-0.037(0.120)	-0.081(0.118)
Bachelor's Degree	0.249(0.118)**	0.264(0.124)**
Professional Degree	0.304(0.218)	0.338(0.208)
Married	-0.121(0.143)	-0.105(0.147)
Wave 1	-0.252(0.138)*	0.0456(0.1248)
Wave 2	-0.168(0.124)	0.0928(0.106)
Wave 3	-0.255(0.1069)**	-0.012(0.093)
Constant	58.074(12.01)**	63.041(10.755)**
Inverse Mills' ratio	2.946(1.369)**	2.839(1.024)**
Adj. R^2	0.3216	0.3188
Fraction due to unobserved component	-	0.472
Number of Observations	9,459	9,459

Table A.3.1. First Stage Results for Weak RE Test using IV. Full Sample

Variables	1st Stage of IV	1st Stage of Corrected IV
<u>Weak RE Test:</u>		
Constant	63.186(0.156)**	105.916(3.942)**
Prob. Of Living to 85	1.120(0.275)**	3.307(0.327)**
Smoking	0.610(0.208)**	0.473(0.212)**
Inverse Mills Ratio	-	-28.471(2.632)**
Economic factors at time t		
Net Worth (in $100,000)	-	0.007(0.018)
Respondent Income (in $1,000)	-	-0.019(0.002)**
No Health Insurance	-	6.937(0.597)**
Private Health Insurance	-	1.377(0.232)**
Self-employed	-	1.575(0.285)**
Pension	-	-7.819(0.626)**
Financially Knowledgeable	-	-0.092(0.201)
Health factors at time t		
Health limitation	-	0.439(0.239)*
Good-V.Good-Exc. Health	-	-1.904(0.326)**
Doctor visits	-	-0.027(0.013)**
High blood pressure	-	-0.284(0.229)
Diabetes	-	1.179(0.406)**
Cancer	-	0.465(1.016)
Stroke	-	-4.496(1.863)**
Heart Problems	-	0.146(0.350)
Arthritis	-	-0.877(0.223)**
Difficulty walking multiple blocks	-	-0.788(0.366)**
Difficulty climbing stairs	-	-0.275(0.472)
Demographic factors at time t		
White	-	0.779(0.215)**
Male	-	-1.142(0.298)**
Bachelor's Degree	-	-0.372(0.225)*
Professional Degree	-	-3.646(0.433)**
Married	-	-1.402(0.235)**
Wave 1-2	-	0.5917(0.2305)**
Wave 2-3	-	-1.7589(0.2603)**
Adj. R^2	0.004	0.097
Test of Weak Instruments	F(2,5021)=12.01	F(29,4601)=15.15
Number of Observations	5,024	4,631

Table A.3.2. First Stage Results for Strong RE Test using IV. Full Sample

Variables	1st Stage of IV	1st Stage of Corrected IV
Strong RE Test:		
Constant	37.959(1.236)**	105.916(3.942)**
Prob. of living to 85	1.191(0.268)**	3.307(0.327)**
Smoking	0.754(0.202)**	0.473(0.212)**
Inverse Mills Ratio	-	-28.471(2.632)**
Economic factors at time t		
Net Worth (in $100,000)	-0.050(0.017)**	0.007(0.018)
Respondent Income (in $1,000)	0.000(0.001)	-0.019(0.002)
No Health Insurance	2.096(0.365)**	6.937(0.597)**
Private Health Insurance	0.340(0.209)	1.377(0.232)**
Self-employed	1.117(0.269)**	1.575(0.285)**
Pension	-1.048(0.201)**	-7.819(0.626)**
Financially Knowledgeable	0.434(0.184)**	-0.092(0.201)**
Health factors at time t		
Health limitation	0.212(0.225)	0.439(0.239)*
Good-V.Good-Exc. Health	-0.018(0.271)	-1.904(0.326)**
Doctor visits	-0.007(0.012)	-0.027(0.013)**
High blood pressure	0.133(0.207)	-0.284(0.229)
Diabetes	0.082(0.374)	1.179(0.406)**
Cancer	-1.252(0.958)	0.465(1.016)
Stroke	-2.638(1.690)	-4.496(1.863)**
Heart Problems	-0.033(0.331)	0.146(0.350)
Arthritis	-0.238(0.193)	-0.877(0.223)**
Difficulty walking multiple blocks	-0.167(0.340)	-0.788(0.366)**
Difficulty climbing stairs	-0.065(0.448)	-0.275(0.472)
Demographic factors at time t		
Age	0.445(0.020)**	-
White	0.481(0.205)**	0.779(0.215)**
Male	0.369(0.179)**	-1.142(0.298)**
Bachelor's Degree	0.018(0.212)	-0.372(0.225)*
Professional Degree	-0.071(0.283)	-3.646(0.433)**
Married	-0.669(0.225)**	-1.402(0.235)**
Wave 1-2	0.5591(0.198)**	0.5917(0.2305)**
Wave 2-3	0.1462(0.2107)	-1.7589(0.2603)**
Adj. R^2	0.161	0.097
Test of Weak Instruments	F(2,4692)=16.01	F(3,4602)=49.18
Number of Observations	4,721	4,631

Table A.4. Tests of Rational Expectations. Restricted Sample

Variables	IV	Corrected IV
Weak RE Test (H_0: Beta =1):	***Cannot Reject***	***Reject***
Constant	-16.283(23.719)	-6.726(2.662)**
Expected Retirement Age$_t$	1.269(0.382)**	1.114(0.043)**
Test of Over-Id Restrictions	Cannot Rej. P-v=.2499	Reject P-v=.0446
Test of Weak Instruments	Cannot Rej. P-v=.086	Reject P-v=.0000
Strong RE Test (H_0: Beta =1):	***Cannot Reject***	***Reject***
Constant	20.089(22.396)	-11.468(7.443)
Expected Retirement Age$_t$	0.371(0.604)	1.200(0.120)**
Economic factors at time t		
Net Worth (in $100,000)	-0.008(0.042)	0.046(0.016)**
Respondent Income (in $1,000)	0.000(0.001)	-0.002(0.001)**
No Health Insurance	0.186(0.315)	-0.008(0.334)
Private Health Insurance	-0.040(0.123)	0.010(0.146)
Self-employed	0.006(0.178)	0.063(0.178)
Pension	-0.228(0.178)	-0.003(0.149)
Financially Knowledgeable	0.187(0.347)	-0.275(0.144)*
Health factors at time t		
Health limitation	-0.011(0.116)	0.032(0.132)
Good-V.Good-Exc. Health	-0.102(0.176)	-0.100(0.207)
Doctor visits	-0.008(0.010)	0.002(0.009)
High blood pressure	0.089(0.140)	-0.031(0.144)
Diabetes	-0.189(0.228)	-0.294(0.280)
Cancer	-0.418(0.372)	-0.142(0.389)
Stroke	-0.062(0.873)	1.128(0.739)
Heart Problems	0.346(0.271)	0.490(0.284)*
Arthritis	-0.185(0.197)	0.031(0.122)
Difficulty walking multiple blocks	-0.168(0.198)	-0.194(0.254)
Difficulty climbing stairs	-0.064(0.223)	0.016(0.250)
Demographic factors at time t		
Age	0.348(0.265)	-
White	-0.261(0.222)	-0.502(0.171)**
Male	0.290(0.181)	0.064(0.163)
Bachelor’s Degree	0.153(0.150)	0.060(0.132)
Professional Degree	-0.105(0.181)	-0.227(0.153)
Married	-0.017(0.167)	0.129(0.170)
Wave 1-2	-0.555(0.182)**	-0.2185(0.148)
Wave 2-3	-0.0429(0.1234)	0.0641(0.1394)
Test of joint Significance of Covariates	Reject. P-v=.0073	Reject. P-v=.0643
Test of Over-Id Restrictions	Cannot Rej. P-v=.8768	Cannot Rej. P-v=.3527
Test of Weak Instruments	Cannot Rej. P-v=.2379	Reject P-v=.0000
Number of Observations	3,905	3,828

Table A.5. New information and plans for retirement. Restricted Sample

Variables	Pooled OLS	IV	Corrected IV
Expected Retirement Age$_t$	0.720(0.031)**	1.066(0.076)**	1.161(0.040)**
Constant	17.824(1.928)**	-3.717(4.713)	-9.567(2.479)**
<u>Economic Factors</u>			
Changes in Net Worth (in 100K)	-0.015(0.013)	-0.008(0.014)	-0.007(0.015)
Changes in Earnings (in $1,000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Changes in Priv. Health Insurance	-0.007(0.114)	0.080(0.134)	0.125(0.140)
Changes in No Health Insurance	-0.320(0.346)	-0.402(0.353)	-0.314(0.375)
Changes in Pension	0.360(0.188)*	0.434(0.198)**	0.406(0.209)*
<u>Health Factors</u>			
Health limitation transitions[36]	-0.086(0.132)	0.003(0.116)	0.011(0.120)
G-V.G.-Exc. Health transitions[37]	0.260(0.181)	0.087(0.178)	0.127(0.189)
Stroke transitions[38]	-0.603(0.404)	-0.920(0.460)**	-0.956(0.487)**
Heart Problems transitions	0.037(0.203)	0.005(0.198)	-0.033(0.208)
Cancer transitions	0.326(0.286)	0.152(0.326)	0.159(0.355)
Diabetes transitions	0.781(0.308)**	0.672(0.343)**	0.682(0.367)*
High blood pressure transitions	0.084(0.103)	0.130(0.112)	0.136(0.117)
Arthritis transitions	0.097(0.080)	0.104(0.090)	0.120(0.095)
Doctor visits	-0.007(0.006)	-0.010(0.007)	-0.011(0.007)
Diff. walking multiple blocks	0.191(0.273)	-0.203(0.269)	-0.200(0.281)
Difficulty climbing stairs	-0.092(0.158)	-0.111(0.212)	-0.099(0.221)
<u>Demographic Factors</u>			
Marriage Transitions[39]	-1.081(0.504)**	-0.729(0.356)**	-0.801(0.404)**
Inverse Mills' ratio First Stage	-	-	45.79(2.21)**
Adj. R^2	0.4119	-	-
Test of Over-Id Restrictions	-	Cannot Reject P-v=.469	Cannot Reject P-v=.564
Test of Weak Instruments	-	Reject P-v=.0000	Reject P-v=.0000
Number of Observations	3,749	3,528	3,457

[36] =1 if new limitation, 0 if no change, and -1 if you got better
[37] positive means health improved
[38] positive means condition (Stroke, cancer, high blood pressure) worsened (0 no change, -1 if better)
[39] =1 means new marriage, = 0 no change, =-1 dissolved marriage (widow, divorce…)