University of Michigan Retirement Research Center

Working Paper
WP 2005-112

On the Covariance Structure of Changes in Consumption in the Health and Retirement Study

Michael Perry

MRRC

Project #: UM05-S1

"On the Covariance Structure of Changes in Consumption in the Health and Retirement Study"

Michael Perry
University of Michigan

January 2006

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104
http://www.mrrc.isr.umich.edu/
(734) 615-0422

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 findings and conclusions expressed are solely those of the author and do not represent the views of the Social Security Administration, any agency of the Federal government, or the Michigan Retirement Research Center.

On the Covariance Structure of Changes in Consumption in the Health and Retirement Study

Michael Perry

Abstract

Using generalized method of moments on the covariance matrix, I test three models of consumption change on constructed consumption data from the Health and Retirement Study. Meant as a first step towards estimating life-cycle effects of subjective survival probabilities on consumption profiles, this study finds that the model that best describes the consumption data is a pure measurement error model. This is likely due to the large amount of error introduced in the process of inferring consumption from other financial data. This result casts significant doubt on the use of this data in estimating a life-cycle model.

(Note: this paper duplicates some expository material from Perry (2005)

I. Introduction

The primary concern of this paper is whether household consumption values constructed from wealth and income data in the Health and Retirement Study are accurate enough to be of use in estimating an Euler equation describing consumption in a model with uncertain lifetimes. This paper is a companion to Perry (2005) in which I actually attempt to estimate that equation.

People's beliefs about their own life-expectancy have not been extensively studied—mainly due to lack of data. It is not clear that people actually have consistent beliefs about their own chances of survival at any time. Even if they do, measuring them in a meaningful and convincing way is difficult.

The life-cycle hypothesis makes a simple prediction about the relationship between a person's perceived risk of death and their consumption: those who think they are less likely to die will have less consumption growth over time. Simply put, if you expect to live a long time, you will conserve your resources early in life in order to have enough later—this means earlier consumption will be lower than it would have been if you had thought your chances of survival were worse, ceteris paribus. In this way, a higher expected chance of survival should have the same effect as a higher interest rate or a lower degree of impatience.

The Health and Retirement Study (HRS) has elicited subjective life-expectation data from its respondents since the study's inception in 1992 (12 waves of the HRS have been completed—1992-2002, every two years). The questions are of the form "What is the percent chance that you will live to be 75 or more?" (the target age—75 in this case—can vary).

The HRS, however, does not elicit consumption data from respondents. Instead, it provides measurements of assets, income and capital gains. These can be used to deduce a consumption level for the time periods between survey interviews. This process leads to a large amount of measurement error, though, as the assets, income and capital gains are all measured with non-trivial measurement error to begin with.

While my primary goal is to study the relationship between consumption profiles and subjective survival beliefs, this study is a first step in which I test constructed HRS consumption data against three alternative models of consumption—specifically, models of the covariance structure of changes in log-consumption. I find that a model describing consumption as an individual-specific constant plus a time-varying random shock best describes the data. This can either be interpreted as showing that the consumption values have too much error to be of significant use for this application or as showing that that process describes real respondent consumption in this dataset. In either case, the result indicates that this dataset is probably not of much use in testing the primary relationship of interest.

In section one I describe the three models of consumption that I test the data against. In section two I describe the data and how I produced my values of consumption. Section three contains the results of fitting the data to the three models.

II. Consumption Models

My method is to propose a model for the consumption data; use it to derive conditions on the covariance between the changes in log-consumption in different periods; and then use those conditions to fit the data to that model using generalized method of moments as described in the appendix of Abowd and Card (1989).

Measurement Error Model

The first model I propose to test is a model in which an individual's log-consumption equals a person-specific constant plus a random shock that changes each period:

$$\ln c_t = c_i + u_t \tag{1}$$

$$U \equiv \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma \right), \Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}.$$

This implies that the change in log consumption between two periods is given by:

$$\ln c_{t+1} - \ln c_t \equiv \Delta c_t = u_{t+1} - u_t , \tag{2}$$

and

$$\operatorname{cov}\left(\Delta c_t, \Delta c_{t+k}\right) = \begin{cases} 2\sigma^2 \text{ if } k = 0 \\ -\sigma^2 \text{ if } k = 1 \\ 0 \quad \text{if k>1} \end{cases} \quad (3).$$

As stated in the introduction, this model can thought of as a model of pure measurement error, or alternatively as actually describing the process that describes real consumption. The only parameter value to be estimated is σ^2.

Random Walk Model

The second model I test is a variation on the random-walk model of consumption proposed by Hall (1978). The variation is that I propose (for convenience) to test whether log-consumption follows a random walk. The proposed model is

$$\ln c_{t+1} = \ln c_t + e_t \tag{4}$$

$$E \equiv \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma \right), \Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}.$$

I ignore the issue of whether there is an additional trend term as that will not be identified by the GMM approach I take. Equation (4) implies

$$\ln c_{t+1} - \ln c_t \equiv \Delta c_t = e_t \qquad (5).$$

Therefore, the covariance structure of changes in log consumption is:

$$\operatorname{cov}\left(\Delta c_t, \Delta c_{t+k}\right) = \begin{cases} \sigma^2 \text{ if } k = 0 \\ 0 \text{ otherwise} \end{cases} \qquad (6).$$

Similar to the measurement error model, the only parameter to be estimated is σ^2. However, this model implies zero covariance for elements that will have non-zero covariance in the measurement error model.

Life-Cycle Model with Uncertain Lifetime

The third model I test is based on the life-cycle implication stated in the introduction. The implication that consumption growth should rise with a rise in a person's mortality risk comes directly from the Euler equation of an agent maximizing the sum of additively separable utility over his lifetime (my formulation is borrowed from Kuehlwein, 1993):

$$E\left(\frac{U'(C_{t+1})}{U'(C_t)} \cdot p \cdot \frac{1+r}{1+\delta} \right) = 1. \tag{7}$$

Here, *p* is the probability of surviving to the next period, *r* is the interest rate and δ is the rate of time-preference. Lowering *p* has the same effect as raising δ or lowering *r*—it privileges current consumption over future consumption. In this way, because I have no measurement of δ, the effect of survival expectations will not be separately identified from the effect of time-preference. I ignore the issue of the utility value of a bequest upon dying.

I assume a felicity function $U(C) = \frac{C^{\gamma}}{\gamma}$ (constant relative risk aversion with relative risk aversion parameter 1-γ), and take the logarithm of each side. Also, because my main concern is with life-span uncertainty, I assume that the income stream is known. This means that there should be no uncertainty about realized consumption in period *t+1*, given that the respondent survives to that period, so I dispense with the expectation operator:

$$\ln c_{t+1} - \ln c_t = \frac{1}{1-\gamma}(\log p_t + \log(1+r_t) - \log(1+\delta)) . \tag{8}$$

Adding a term to account for measurement error in the change in log-consumption gives:

$$\Delta c_t = \frac{1}{1-\gamma}(\log p_t + \log(1+r_t) - \log(1+\delta)) + u_t . \tag{9}$$

I assume that $E(u_t) = 0$ and $\text{var}(u_t) = \sigma^2$ for all *t*.

As written, equation (9) should apply only to a single agent making decisions for himself. Analyzing a similar case for a multi-person household in which agents care for each other’s well-being requires further assumptions about how those agents interact and make decisions together.

The only variables from (9) that I have measured variation in are consumption and subjective survival expectation. Therefore, I make the possibly unfounded assumptions that the difference between *log* r_t and $log\,\delta$ is distributed randomly in the population given $log(p_t)$, and that γ is constant (or distributed randomly) throughout the population. This leaves the relationship that I examine:

$$\Delta c_t = K + \beta \log p_t + u_t\,, \tag{10}$$

substituting $\beta = \dfrac{1}{1-\gamma}$ and $K = \dfrac{1}{1-\gamma}\log\left(\dfrac{1+r}{1+\delta}\right)$. This equation implies the covariance structure:

$$\operatorname{cov}\left(\Delta c_t, \Delta c_{t+k}\right) = \begin{cases} \beta^2 \operatorname{var}\left(\log p_t\right) + \sigma^2 & \text{if } k = 0 \\ \beta^2 \operatorname{cov}\left(\log p_t, \log p_{t+k}\right) & \text{otherwise} \end{cases} \qquad (11).$$

The GMM estimation of this model will fit values of both σ^2 and β^2. Note that if β^2=0, then this reduces to the random-walk model.

III. Data

The HRS is a nationally representative panel study of persons over 50 in the United States. Beginning in 1992, respondents were interviewed every two years, covering health, finances, physical and mental capabilities, family structure and relationships and job history. A study called AHEAD (Assets and Health Dynamics of the Oldest Old) began in 1993 and focused on older respondents. In 1996, the AHEAD study merged with HRS. New cohorts were added to HRS in 1998 so that the survey would remain representative of those over 50. The last wave of data available for this analysis comes from interviews done in 2002. I employ all HRS waves, but I do not use AHEAD data that was taken prior to the merger with HRS.

Table 1 shows descriptive statistics of the population that I will use for this analysis. In each survey wave this population consists of all respondents who answered at least one subjective survival question in that wave and who were single throughout the time 1992-2002. P(75) and P(85) refer to the mean values of the probability responses to the subjective survival questions that ask about target ages of 75 and 85 respectively[1]. These statistics are meant simply to make it clear what the population of respondents is like in any particular year. They cannot be used to make accurate inferences about the evolution of households or singles in the HRS population over time because those respondents who have a valid answer to at least one subjective survival question are a highly non-random group. This is due to both self-selection (it takes a certain mental capacity to give a sensible answer to a probability question) and due to survey variation (exactly which sets of respondents have been asked which questions has varied over time in the HRS).

Measuring Consumption

Each wave of the HRS contains detailed questions on household assets (both real and financial), household income (separate from capital gains), and capital gains. The survey does not contain any consistent measure of household consumption. In order to test the implication of survival expectations on consumption profiles, I use the HRS data on assets, income and capital gains to infer a measure of consumption for each respondent for each period between survey interviews.

[1] P(85) is a misnomer for the 2000 and 2002 waves because in those waves the target age of the probability question varied based on respondent age. This is described in detail later.

The basis of the calculation is the relationship:

$$C = I + CG - \Delta A. \tag{12}$$

That is, consumption between two measured points in time equals whatever the household took in, in earned income and capital gains, minus the amount that their asset level grew during that period. It is ambiguous in the HRS whether respondents give pre-tax or post-tax income levels and so there is no way to account for income tax. I do, however subtract property taxes from inferred consumption.

First, I use the HRS income data to estimate household income over the period between survey interviews. I divide the study's constructed household income variable—which estimates total household income in the one-year period prior to the interview—by twelve to get an estimated monthly income and then multiply by the number of months between interviews. This procedure will add measurement error to the extent that actual household income during the period between interviews differs from income during the period just prior to the interview. Additionally, all financial variables in the HRS include imputed values which increase the level of measurement error, but also substantially increase the number of data points available. To exclude the imputed values from this analysis would entail dropping a majority of the available data since almost all respondents require imputation on at least some financial variables.

Second, I use the capital gains section of the survey to estimate capital gains between survey interviews. Respondents are asked whether they have put money in to or taken money out from their various assets. This information, combined with the asset values reported in the earlier and later waves, allows for inference of the respondent's capital gains over the period. This is straightforward except that housing capital gains are not well-measured for respondents who buy or sell a house during the period, so those respondents are dropped.

Finally, I calculate respondents' change in assets between the survey interviews by subtracting the later survey-interview household assets variable from the earlier survey-interview household assets variable. I do not include housing assets on the assumption that people—particularly retired people—do not generally monetize housing assets for the sake of consumption. Descriptively, this assumption is probably alright for the period 1992-2002, but may be less so now.

Adding income and capital gains and subtracting asset growth and property taxes and then deflating by the CPI-U yields the measure of consumption in 2002 dollars that is used to test the life-cycle prediction.

Using this strategy I have measures of consumption for the periods between the survey waves 1992 and 1994, 1994 and 1996, 1996 and 1998, 1998 and 2000 and 2000 and 2002. These five sets of consumption data can be used to calculate four cross-sections of log-consumption growth, the statistic of interest in the Euler equation. Table 2 shows summary statistics for the measures of consumption in 2002 dollars and log-consumption growth. Panel A shows consumption measured for all households consumption measured for households composed of singles. Panel B shows log-consumption growth measured for all households and for singles. The values in Panel A can be thought of as approximately 2-year levels of consumption for those households as that is approximately the time between survey interviews. Panel C, most relevant for this study, shows consumption and log-consumption growth for those singles who exist throughout the entire panel period and who have no negative consumption values—these are the respondents the model is estimated for. This group is different from the overall singles group in two major ways. First, consumption levels are much more steady through the period—perhaps because it is always the same group of people. Consumption for each approximately two-year period is always approximately $50-60K. Second, for both of the other groups (all households and all singles) consumption growth is negative for the first three periods and positive in the last. For this group of singles, consumption growth is negative in the first period and positive thereafter. I have no explanation for that difference.

Two elements of this table suggest large measurement error. First, in each year a large proportion of households have negative values for this measure of consumption—typically 11-15%. Because actual consumption cannot be negative, these cases are necessarily mis-measured. The proportion of negative cases is a lower bound on the proportion of mis-measured cases in each year. The large number of negative values also explains why the number of cases is significantly lower in Panel B and Panel C than Panel A—if there is a negative value in either the early period or the late period, then log-consumption growth cannot be measured.

Second is the fact that, for both the whole population and for singles, measured consumption drops substantially for the period 1998-2000 and then rises substantially for the period 2000-2002. This may be due to unreported capital gains appearing in the change in asset level. If a respondent had substantial capital gains in 1998-2000 (as many did), then did not report them as capital gains and did correctly report their total assets, this would result in measured consumption being biased downwards. The reverse is likely for the period 2000-2002. Poorly measured capital gains are probably not restricted to these time periods—they just show up strongly in these periods because asset values fluctuated substantially.

The HRS does provide a few variables that can be used to corroborate my deduced consumption values. In 1996 and 1998 the survey asked each household what their total spending—including all debt payments, utility bills, rent, transportation, entertainment, food, clothes and any other expenses—was in the previous month. Also, in 2002, the survey asked three food consumption questions: how much did the household spend in the past week on all food; how much did it spend having food delivered; and how much did it spend eating out. The left side of table 3 shows mean values for these measures and for deduced consumption values measured on a monthly basis for the same time period. The HRS survey levels of consumption are substantially lower in both cases. Also included are mean values of inferred consumption with negative values removed—this exacerbates the difference between the HRS measure and my measure. It is questionable how accurate a respondent is likely to be in making a fast estimate of monthly spending, so there is no guarantee that the HRS measure is very good. One possible, partial explanation for the large difference between the values reported by respondents and the calculated values is that I have not accounted for income tax. If respondents generally report pre-tax income, then my calculation will count their taxes as consumption. It seems likely that few respondents would include income tax in their response to the 1996 and 1998 HRS consumption question.

The right side of table 3 shows correlation coefficients and respective significance levels between the HRS measures and my inferred levels of consumption for the relevant time periods. For both HRS consumption measures, the correlation is substantially higher when the negative cases are removed from the deduced consumption numbers. This is unsurprising as those cases almost certainly represent particularly egregious cases of measurement error. Furthermore, it is encouraging that the inferred consumption shows such a high correlation with the HRS measures of consumption given the likely presence of substantial measurement error in both. Interestingly, of the food consumption measures in 2002, only the measurement of what a family spends eating at restaurants is significantly correlated with my inferred consumption measure. Also interesting, though I have not shown it, is that these measures of food consumption correlate very little with each other—again probably due to measurement error.

Table 3 shows that despite its weaknesses, my inference about household consumption do match up to a substantial degree with the limited information the HRS survey provides about actual household consumption.

Measuring Subjective Survival Expectations

In the Euler equation that provides the hypothesis tested, p represents the agent's subjective assessment of his probability of living to the next period. The HRS provides answers to questions of the form "What is the percent chance that you will live to be 75 or more?" These questions are asked twice in each survey wave with different target ages, although some respondents may only be asked once or not at all. From 1992 to 1998 respondents were asked the questions with 75 as a target age and then with 85 as a target age. In 2000 and 2002, the first question remains the same and the second question has a target age that varies from 80 to 100 in five year increments depending on the age of the respondent (the target for anyone under 70 was 80, for those 70-74 it was 85 and so on).

A response to one of these questions does not imply directly any particular value of the respondent's expected chance of living to any particular date other than the target age. In order to use the survey responses to calculate a value of *p* (in the life-cycle model the probability of living to the next period; in this analysis the probability of living through the next period of measured consumption) for each respondent, some assumptions are necessary. In the appendix, I describe an algorithm I use to produce values of *p* from the responses to the subjective survival questions in the survey.

Covariances of Log-Consumption Growth

The object I use to fit the three proposed models to the data is the covariance matrix of log-consumption growth. This is shown, along with the corresponding correlation matrix in table 4. Significance levels are shown for the correlation coefficients as well (a value of 1%, e.g., indicates that the correlation is very significant). As can be shown relatively easily, the pure measurement error model implies a correlation matrix with -1/2 on the elements one-removed from the diagonal in the correlation matrix and zero for the other off-diagonal elements. Prima facie, the correlation matrix of the data appears very close to that—with one-off diagonal elements close to -1/2 and highly significant and no other significant correlations.

IV. Results

The results of GMM estimation are presented in table 5. Results are shown both for optimal minimum-distance (OMD) estimation, using the inverse of the variance matrix of the vector of covariance elements as a weighting matrix, and for equal-weighted minimum distance (EWMD) estimation, using the identity matrix as a weighting matrix.

In both cases, we cannot reject the measurement error model at standard significance levels and we can reject the other two models. Indeed, the random-walk model and the life-cycle model both fit the data very poorly—producing very high chi-square statistics. (Note in table 5, I use the term p-value to indicate the probability of observing a χ^2 value at least that high given the proposed model.)

There are two other results to note in table 5. First, the value of β^2 that minimized the test statistic for the life-cycle model is zero. This reduces the life-cycle model to the random walk model—indeed the values for σ^2 are the same for each model. The models produce different χ^2 values in the EWMD case because although they minimize the same statistic, the test statistic for that minimization depends on the first derivative of the vector of covariance elements with respect to the parameter vector. In the life-cycle model this derivative depends on the covariances of p (the subjective survival probabilities). In the random walk model, this derivative is a vector of constants.

Second, OMD and EWMD produce very similar parameter values for the well-fitting model, but very different values for the poorly-fitting model. Abowd and Card note this issue and that is their justification for using both—it serves as another test of how well the model fits.

The primary implication of these results is that this data is unlikely to be of much use in estimating the life-cycle model that is really the object of interest. Because of the procedure used to produce the consumption data, I conjecture that the results of this study indicate a large amount of measurement error in the data, rather than a real consumption process for survey respondents.

As the main goal of this research program is to investigate whatever link may exist between people's stated survival beliefs and their decision-making, it will be necessary either to derive a testable result that does not rely on consumption data or to find a dataset that measures both survival expectations and consumption.

References

Abowd, J.M., and Card, D. (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411-445.

Hall, R. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *The Journal of Political Economy*, 86, 971-987.

Kuehlwein, M. (1993): "Life-Cycle and Altruistic Theories of Saving with Lifetime Uncertainty," *The Review of Economics and Statistics*, 75, 38-47.

Perry, M. (2005): "Estimating Life-Cycle Effects of Survival Probabilities in the Health and Retirement Study," *Michigan Retirement Research Center Working Paper*.

Appendix: Calculating Values of *p* from Survey Responses

For each respondent, I assume that a response to the question "What is the percent chance that you will live to be 75 or more?" implies a belief over all the conditional probabilities of surviving one year into the future (that is, for example, the probability of surviving to age 63 given that the respondent has survived to age 62) for each year from the respondent's current age up to the age of 75. Assuming that these probabilities exist, the response to the question is just the product of all conditional probabilities from the respondent's age to the target age:

$$R = \prod_{i=A+1}^{i=T} \rho_i \,. \tag{13}$$

Here, R is the survey response, A is the age of the respondent, T is the target age and ρ_i is the probability of surviving to age i, given that the respondent has survived to age i-1. In order to calculate values of ρ_i, it is necessary to assume something about how respondents' beliefs change over time. I assume that respondents recognize that their conditional survival probabilities fall somewhat each year that they age[2]. Over the relatively short time of a decade, actual life-table survival probabilities decline approximately linearly. For this reason and for simplicity, I assume that ρ_i declines by a constant amount each year. This simplifies the expression in (13) to

$$R = \prod_{i=1}^{i=T-A} (\rho_{A+1} - (i-1)m) \,, \tag{14}$$

where m is the amount by which survival probabilities decrease each year and H=T-A. Taking logs of both sides gives

$$\ln R = \sum_{i=1}^{i=H} \ln(\rho_{A+1} - (i-1)m) \,. \tag{15}$$

Then, set $\rho_{A+1} = 1 - r$. Actual mortality rates, even for people into their eighties are typically below 0.1—meaning that in actual outcome, survival probabilities are quite

[2] This assumption may be reasonable for those respondents who gain no new and significant information about their life-expectancy during the relevant time period. It is almost certainly not reasonable for respondents who do receive such information by, for example, suffering a major health shock such as a stroke.

close to one for any given year. I assume that respondents' beliefs conform well-enough to actual outcomes that I can use the approximation $\ln(1-x) \cong -x$ in equation (15). This yields

$$\ln R = \sum_{i=1}^{i=H} (-r-(i-1)m) \tag{16}$$

$$\ln R = -rH - m\frac{(H-1)(H)}{2}, \tag{17}$$

where equation (17) follows by using standard summation results.

This is the equation I use to describe the relationship between a response and a respondent's beliefs in the first wave of the survey (1992). Because other responses occur at different times and for different target ages, there is variation in the values of *H* and therefore the multipliers of *r* and *m*. For example, if equation (17) represents the relationship between beliefs and response for the question with target age 75, asked in 1992, then the same relationship for target age 80, asked in 2000 of the same respondent looks like:

$$\ln R_{2000,80} = -rH_{2000,80} - m\left(\frac{(H_{2000,80}-1)(H_{2000,80})}{2} + 8\right) \tag{18}$$

where $H_{2000,80} = H - 8 + 5 = H - 3$. The difference in *H* is due to the respondent's age having advanced eight years between surveys and the target age increasing by five years. The addition of 8 to the multiplier on *m* is due to all conditional survival probabilities having declined by 8*m* as the respondent aged during the time between 1992 and 2000.

Using these relationships, I have a vector of responses, *R*, and a matrix of multipliers for *r* and *m*, *X*, for each respondent. This allows me to estimate the regression

$$\ln R = X' \begin{pmatrix} r \\ m \end{pmatrix} + \varepsilon \tag{19}$$

separately for each respondent, thereby giving values of *r* and *m* for each respondent who has answered at least three subjective survival questions. In this formulation, *r* is the respondent's perceived risk of death in the first year (set to 1992 for all respondents since that is the first year of survey data for any respondents), and, again, *m* is the yearly increase in risk of death. Using these numbers, I calculate a respondent's perceived risk of death in year *x* as $risk_x=r+(x\text{-}1992)m$, or equivalently, I calculate their perceived probability of survival during year *x* as $p_x=1\text{-}risk_x$.

The above explanation elides the issue that when a respondent answers zero, it is impossible to take a log and use that response in the calculation. I implement two strategies to deal with this issue and test which seems to work better. First, I exclude all responses of zero from the calculations. Above, we established that those who answer 100 seem to be similar to those who answer zero, and additionally, they seem to be the same sort of unlikely answer to a probability question—perhaps due to misunderstanding. For that reason, when I exclude the zeros I also exclude the 100s. For the second strategy, instead of excluding the zeros and 100s, I replace the zeroes with the value 0.00001, which can be logged, and I replace the 100s (really 1s since everything is converted to fractions) with 0.99999.

In each case, I use the values of ρ' and *m* generated for each respondent to calculate the respondent's perceived probability of survival during any year. These predicted yearly subjective survival values can be multiplied together as in (13) to produce predicted responses to any of the subjective survival questions on the survey. To test my two strategies, I regress the actual responses on the predicted responses. The results are shown in Appendix Table 1. Strategy 1 drops responses of zero or 100, strategy 2 replaces them. The first two sets of R^2 values are for regressions over the same responses. The last set is for strategy 2 used to predict for all responses for which it is possible to do so. The number of possible cases using strategy 2 is larger because dropping responses in strategy 1 necessarily means reducing some respondents to below the three-response level necessary for prediction. Strategy 1 produces a better set of predicted responses in all cases. This could be because answers of zero or 100 are more likely to reflect confusion than information about held beliefs.

Perhaps needless to say, I do not hypothesize that any respondent has set beliefs about his or her conditional probability of surviving during any particular year. It would be claiming too much to say that the HRS questions evoke anything more than a general impression of survival probability from most respondents (the exception perhaps being any professional actuaries surveyed). The scheme I propose for integrating all of a respondent's answers is intended to be a fairly straightforward way of approximating what a respondent's well-articulated beliefs might look like if they were forced to develop them in a rigorous way and if they had some consistency over time. Therefore, the charge could easily be leveled that I have invented an index with a dubious epistemic nature. My only response is that I see no other simple strategy for incorporating all of a respondent's answers that is not at least as questionable. It may well be that questions like those on the HRS are simply not sophisticated enough to use in testing life-cycle models.

Table 1: Descriptive Statistics for Single HRS Respondents used in this study

	1992	1994	1996	1998	2000	2002
N	599	599	599	599	599	599
mean age	57	59	61	63	65	67
%male	23	23	23	23	23	23
%white	55	55	55	55	55	55
mean p(75)	64	64	64	66	65	68
mean p(85)	44	46	49	50	54	57
mean assets (2002 $K)	45	50	51	45	47	47
mean annual income (2002 $K)	26	25	28	29	27	28

Table 2: Measured Consumption and Log-Consumption Growth

A: Consumption

All Households	N	mean (2002 $K)	stand. dev. (2002 $K)	%negative
1992-1994	6888	81	242	15
1994-1996	6343	82	276	13
1996-1998	6153	91	463	12
1998-2000	12565	69	393	13
2000-2002	11532	105	338	11

Singles	N	mean	stand. dev.	%negative
1992-1994	2040	33	121	19
1994-1996	1952	46	295	15
1996-1998	2038	59	311	12
1998-2000	5698	37	273	14
2000-2002	5506	66	301	13

B: Log-Consumption Growth

All Households	N	mean	stand. dev.
92/94-94/96	4586	-0.004	1.26
94/96-96/98	4430	-0.070	1.23
96/98-98/00	4224	-0.055	1.24
98/00-00/02	8298	0.060	1.25

Singles	N	mean	stand. dev.
92/94-94/96	1188	-0.075	1.29
94/96-96/98	1282	-0.021	1.25
96/98-98/00	1327	-0.045	1.21
98/00-00/02	3560	0.049	1.22

C: Consumption and Growth for those Not Excluded from Model

Consumption

Singles	N	mean	stand. dev.	%negative
1992-1994	599	55	67	0
1994-1996	599	50	63	0
1996-1998	599	53	78	0
1998-2000	599	51	65	0
2000-2002	599	61	110	0

Log Consumption Growth

Singles	N	mean	stand. dev.
92/94-94/96	599	-0.10	1.22
94/96-96/98	599	0.20	1.14
96/98-98/00	599	0.23	1.13
98/00-00/02	599	0.05	1.26

Table 3: Comparison of HRS Consumption Measures with Inferred Consumption

	mean (2002 dollars)		correlations	
1996 (survey)	1722		1996 (inferred)	1996 (inferred, ≥0)
1998 (survey)	1959	1996 (survey)	0.14	0.30
1996 (inferred)	3682			
1998 (inferred)	4398		1998 (inferred)	1998 (inferred, ≥0)
1996 (inferred, ≥0)	5438	1998 (survey)	0	0.33
1998 (inferred, ≥0)	6382			
			2002 (inferred)	2002 (inferred, ≥0)
All food (2002 weekly)	86	All food	0	0
Restaurants	26	Restaurants	0.14	0.18
Delivered	1	Delivered	0	0

Table 4: Covariance and Correlation of Log-Consumption Growth

Covariance Matrix of Change in Log-Consumption

	1	2	3	4
1	1.50			
2	-0.73	1.31		
3	0.08	-0.62	1.27	
4	-0.05	0.01	-0.75	1.59

Correlation Matrix of Change in Log-Consumption (significance level in parentheses)

time period	1	2	3	4
1	1.00			
2	-0.52	1.00		
	(<0.01)			
3	0.05	-0.48	1.00	
	(0.20)	(<0.01)		
4	-0.03	0.00	-0.53	1.00
	(0.56)	(0.92)	(<0.01)	

Table 5: Estimation Results

Results of OMD GMM Estimation				
model	parameter values	Chi-square DOF	Chi-Square Statistic	P-value
Measurement Error	σ-squared=0.684	9	7.67	56.8%
Random Walk	σ-squared=0.198	9	189.33	0.0%
Life-Cycle	β-squared=0 , σ-squared=0.198	8	189.33	0.0%

Results of EWMD GMM Estimation		Chi-square DOF	Chi-Square Statistic	P-value
Measurement Error	σ-squared=0.643	9	7.41	59.5%
Random Walk	σ-squared=1.42	9	189.33	0.0%
Life-Cycle	β-squared=0 , σ-squared=1.42	8	189.33	0.0%

Appendix Table 1: R-squared values for Regressions of Actual on Predicted Responses

		Strategy 1	Strategy 2		Strategy 2 on all eligible cases	
		R2	R2	N	R2	N
1992	P(75)	69	46	6396	49	8310
	P(85)	50	35	6561	51	8382
1994	P(75)	60	39	6390	44	8446
	P(85)	55	38	6828	49	8141
1996	P(75)	51	35	5601	44	8074
	P(85)	61	44	6369	52	7879
1998	P(75)	50	36	6139	41	8320
	P(85)	70	52	6246	56	7400
2000	P(75)	50	35	5761	41	7585
	P(85)	45	37	7314	44	9372
2002	P(75)	44	31	4776	38	6288
	P(85)	54	37	7534	45	9086