

NBER WORKING PAPER SERIES

INDIVIDUAL SUBJECTIVE SURVIVAL CURVES

Li Gan
Michael Hurd
Daniel McFadden

Working Paper 9480
<http://www.nber.org/papers/w9480>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2003

The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

©2003 by Li Gan, Michael Hurd, and Daniel McFadden. All rights reserved. Short sections of text not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including notice, is given to the source.

Individual Subjective Survival Curves
Li Gan, Michael Hurd, and Daniel McFadden
NBER Working Paper No. 9480
January 2003
JEL No. C81

ABSTRACT

Testing life-cycle models and other economic models of saving and consumption at micro level requires knowledge of individuals' subjective beliefs of their mortality risk. Previous studies have shown that individual responses on subjective survival probabilities are generally consistent with life tables. However, survey responses suffer serious problems caused by focal responses of zero and one. This paper suggests using a Bayesian update model that accounts for the problems encountered in focal responses. We also propose models that help us to identify how much each individual deviates from life table in her subjective belief. The resulting individual subjective survival curves have considerable variations and are readily applicable in testing economic models that require individual subjective life expectancies.

Li Gan
The University of Texas at Austin
Department of Economics
1 University Station C3100
BRB 1.116
Austin, TX 78712
gan@eco.utexas.edu

Michael Hurd
RAND Corporation
1700 Main Street
Santa Monica, CA 90407
and NBER
mhurd@rand.org

Daniel McFadden
University of California
Department of Economics
549 Evans Hall #3880
Berkeley, CA 94707-3880
and NBER
mcfadden@econ.berkeley.edu

I. Introduction

Many economic models are based on the forward-looking behavior of economic agents. Although it is often said that “expectations” about future events are important in these models, it is the probability distributions of future events that influence the models. For example, an individual’s consumption and saving decisions are believed to depend upon concerns regarding future interest rates, the likelihood of dying, and the risk of substantial future medical expenditures. According to our theories, decision makers have subjective probability distributions about these and other events in their lives, and moreover, use them to make decisions about their saving practices.

A typical objective of empirical models on intertemporal decision making is to estimate responses to changes in variable levels, such as changes in saving due to an anticipated change in the interest rate. A second objective is to find the extent of an individual’s risk aversion; namely, what is his or her response to changes in outcome variability? For instance, do changes in the variability of future income lead to changes in saving practices? These are worthwhile objectives due to the importance of society of choices that depend on uncertain future events. For example, poverty in old age depends partly on an individual’s consumption choices at a younger age. Consequently, how is consumption influenced by mortality risk and the uncertainty of medical expenditures? Why do some individuals purchase adequate insurance against unfavorable outcomes while others do not? Why do many reach retirement age with inadequate financial provisions for post-retirement living expenses? Is it due to misperceptions about the probabilities of reaching old age? Do people maintain excessive housing into old age as a

hedge against inflation risk? The answers to these and similar questions depend on our understandings of decision maker reactions to future uncertainty. Moreover, creating policies that alleviate the consequences of such decision-making processes depends on answers to the aforementioned questions.

In a few economic models, we have data on probability distributions that are assumed to approximate those required by decision making models under uncertainty. Life cycle models of consumption, in which mortality risk helps determine savings, have been estimated by assuming that individuals have subjective probability distributions on mortality risk that are the same as those found from life tables (Hurd 1989). A precautionary motive for saving thus depends on the risk of future medical outlays. It therefore seems reasonable that the distribution of outlays as estimated from micro data represents a good approximation of the subjective probability distributions used by decision makers (Hubbard, Skinner and Zeldes, 1995). More generally, Manski (1993) has proposed using observed outcome probabilities in panel data as estimations of the subjective probability distributions for individuals on the panel on the grounds that the sampling exercise can itself be taken as a “model” of the subjective probability process.

In most applications, however, we do not have adequate data for probability distributions thus requiring the use of unverifiable assumptions in estimations. For example, in macroeconomic models expectations are assumed to be rational, which often yields an estimated relationship. Yet, the rationality assumption cannot be tested outside of the model’s immediate context. In life cycle models on saving, a cohort’s average mortality risk may not be well approximated by the mortality risk found on life tables due to changes in risk; that is, a cohort may not believe that the mortality experience of older

cohorts will be the same as his or her own. Furthermore, individuals within the same cohort will have different subjective evaluations of probability distributions and its influence on their behavior, even if it is systematically incorrect. However, such evaluations are not generally observable. These individual heterogeneities often become problematic in parameter estimates. For example, consider a typical individual utility function,

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma},$$

where c_t is the consumption at time t , and γ is the risk aversion parameter. The first order condition in a common formulation is:

$$\frac{1}{c_t} \cdot \frac{dc_t}{dt} = \frac{-h_t + r - \rho}{\gamma} + \beta' X_t,$$

where h_t is the individual subjective hazard rates, while X represents certain socio-demographic variables. In this framework, if h_t is not observed but correlated with X_t , we will have a typical endogeneity problem. If h_t is poorly measured, estimations of γ will subsequently be biased.

Previous studies have typically obtained individual mortality risks through two different approaches: either by using life tables or by using well-known variations in mortality rates by economic status. Since mortality risk life tables only vary by age, race, and sex, there are not enough variations from which to calculate mortality risks. If subjective mortality risks of individuals with different economic status vary in the same way as observed mortality rates, model estimations using standard life tables will lead to biased estimates. Moreover, forecasts of economic status distributions will be incorrect such that poorer individuals who believe that their mortality risk is higher will spend

money faster than what is predicted by the model. Yet, according to observable characteristics, mortality risk variations can, in principle, be calculated given that some economic variables are potentially endogenous. More importantly, individuals surely have subjective probability distributions that are partly related to observable variables.

Two recent surveys have posed questions regarding individual subjective probabilities including, Asset and Health Dynamics among the Oldest Old (AHEAD) and the Health and Retirement Survey (HRS). Hurd and McGarry (1995) reveal that average survival probabilities are very close to those presented in life tables. In a more recent paper, Hurd and McGarry (2002) use panel data from HRS and find that respondents modify their probabilities in response to new information such as the onset of a new illness. Their findings are consistent with an earlier study of Hamermesh (1985) who surveys a selected sample of economists about their survival probabilities. Although on average self-reported survival probabilities are consistent with life tables, at the personal level, however, these probabilities face a serious problem. In all age groups, we find that a large fraction of respondents give what we call focal-point responses: 0.0 and 1.0. These responses cannot represent the respondents' true probabilities as the distribution of true probabilities should be continuous, and moreover, true probabilities cannot literally equal zero or one. Thus the main focus of this paper is to "recover" the "true" subjective survival curve for each respondent. To do so, we develop a Bayesian update model to accomplish this objective.

In our model, for individuals at age a , we let the prior survival probability distribution at a future point in time ($a+t$) be a truncated normal between zero and one (we do not include zero and one). The conditional density of the observed survival

probability is assumed to be a censored normal between zero and one, allowing for the focal points. In addition, we suggest two approaches that model the deviations of each individual's belief from the life table.

We use the posterior density mean as an individual's estimated subjective survival probabilities, and estimate the model using the observed death record. Our model produces optimistic indices to measure the deviation of her subjective belief from the life table. Consequently, the survival curves for each individual produced by the optimistic indices do not encounter problems associated with focal points and have considerable variations. These subjective survival curves are readily applicable to life cycle models and other economic models that require individual subjective mortality risk.

The remainder of the paper is organized as follows: Section 2 introduces the self-reported subjective survival probabilities including their consistency with the life table and problems associated with individual responses. Next, Section 3 introduces a Bayesian method that helps us to recover underlying subjective survival curves. Section 3 also introduces two approaches that are used to represent individual deviations from life tables. In Section 4, we estimate the model and conduct the out-of-sample prediction. Lastly, we present the paper's conclusions in Section 5.

II. Individual Subjective Mortality Risk

In the AHEAD sample, each respondent is asked a series of questions about how likely various presented future events will occur. These future events include: an income that is consistent with changes in inflation, major medical expenses, leaving a bequest, receiving financial help from family members, moving to a nursing home, and surviving

for another 10-14 years.² In particular, the survival probability question AHEAD posed to respondents is as follows:

[Using any] number from 0 to 100 where “0” means that you think there is absolutely no chance and “100” means that you think the event is absolutely sure to happen ... What do you think are chances that: You will live to at least A?(A is an age that is 11-15 years older than the respondent’s current age)

To examine whether these survival probabilities carry useful information, we compare the subjective survival probabilities with the life tables. Table 1 lists the average and median survival probabilities from AHEAD and the 1992 life tables for the target ages used in the AHEAD survival questions as calculated by the first two waves of AHEAD (e.g., 85 years of age for subjects aged 70-74, 90 years of age for subjects aged 75-79, etc.). In general, younger AHEAD respondents have average subjective probabilities that closely mirror life table averages, while older respondents have averages that are substantially higher.³ In general, AHEAD medians are closely related to those in the life table.

Table 2 lists the percentage of those respondents who gave continuous responses, focal responses, and no responses in the two waves. Table 2 also lists the transition probabilities of different response modes between the two waves. In wave 1, only 41.5% of respondents gave continuous responses, with more than 30% of them providing either

² Bassett and Lumsdaine (2001) find that all responses contain a common component.

³ Several reasons are suggested in Hurd, McFadden and Gan (1997) for this finding. One reason is that the AHEAD survey does not include respondents who reside in nursing homes or other institutional care facilities. Thus AHEAD represents a healthier population than is represented by a life table.

zero or one as their answers. The subjective probabilities for the remainder of the population are not available. In wave 2, more than 75% of respondents gave continuous responses, where approximately 19% of the population either responded zero or one. Thus the prevalence of focal-point responses indicate that subjective probability measurements in AHEAD cannot represent the respondents' true probabilities. Without correcting for focal responses of zero or one, it is impossible to derive a survival curve that varies over time. Thus the primary objective of this paper is to “recover” the “true” subjective survival curve for each respondent, especially for those who give focal responses of zero or one. In the next section, we develop a Bayesian update model to achieve this objective.

III. Modeling Individual Subjective Survival Curves

Before we present the model, it is necessary to define the notations that we use throughout this paper.

- a : age.
- t : time at risk.
- $L_o(t)$: life table survival probability from birth.
- $S_{oa}(t) = L_o(a+t) / L_o(a)$: life table survival probability from age a .
- $\Lambda_\theta(t)$: life table integrated mortality hazard rate.
- $\lambda_\theta(t)$: life table mortality hazard rate.
- T : an age at which $L_o(T) = 0$, say $T = 108$.
- i : individual.

- $S_{ia}(t)$: personal survival probability from age a to target age $a + t$ for subject i .
Since survival probabilities differ for different people at the same age a , we let $S_{ia}(t)$ be a random variable with a density $\pi(s_{ia}(t))$, or $\pi(s_{iat})$.
- $\Lambda_{ia}(t)$: personal integrated mortality hazard rate at age a .
- $\lambda_{ia}(t)$: personal mortality hazard rate at age a .
- τ : time at risk in interview survival question.
- p_{iat} : response to interview survival question. We assume that p_{iat} is measured with an error. The density of p_{iat} conditional on personal survival probability from age a to age $a + t$ is given by: $f(p_{iat} | S_{iat} = s_{iat})$.

By definition, an individual i 's survival curve is:

$$S_{ia}(t) = \exp(-\Lambda_{ia}(a+t) + \Lambda_{ia}(a)) = \exp\left(-\int_0^t \lambda_{ia}(a+r)dr\right) \quad (1)$$

It is first necessary to specify the plausible families of $\lambda_{ia}(a+t)$ that satisfy this equation. We propose to use the population hazard function $\lambda_{0a}(a+t)$ as a base, while minimally modifying it to calculate individual $\lambda_{ia}(a+t)$. Two alternative ways to specify the $\lambda_{ia}(a+t)$ function include:

$$\lambda_{ia}(a+t) = \gamma_i \lambda_{0a}(a+t). \quad (2)$$

The parameter γ_i is an individual “optimism” parameter. In comparison with the life table, if $\gamma_i > 1$, then the person is “pessimistic”; however, if $\gamma_i < 1$, then the person is “optimistic”. Since this model in (2) scales the population hazard, we will refer to it as a “hazard-scaling” model from now on.

The second model specification is given as:

$$\lambda_{ia}(a+t) = \lambda_{0a} \left(a + \frac{t}{\gamma_{ia}} \right) \frac{1}{\gamma_{ia}}. \quad (3)$$

This model represents an accelerated failure time frame where the individual thinks of herself as aging forward from her current age more or less rapidly than the average person. If a large γ_i corresponds to slow future aging, i.e., $\gamma_i > 1$, then the person is “optimistic;” however, if $\gamma_{ia} < 1$, then the person is “pessimistic”. Similarly, we refer to the model in (3) as the “age-scaling” model as it scales ages to represent individual optimism.

If there is no response error or focal bias in p_{iat} , the models in (2) and (3) are accurately identified having no free parameters. We can then take these models as actual survival information, and subsequently, decide which model works best. If a response error or focal bias in p_{iat} is present, then the personal survival curve is not forced through p_{iat} at age $a + \tau$. In this case, we use a Bayesian framework instead. Our basic assumptions are addressed below. Prior belief for the personal survival curve density S_{iat} is $\pi(s_{iat})$. The mean for prior density is $\exp(-\psi \Delta \Lambda_{0at})$ where ψ represents a parameter for measuring the population’s average subjective optimistic degree (when $\psi = 1$, the mean of prior distribution S_{iat} corresponds with the life table value). Given S_{iat} , the self-reported survival probability p_{iat} has a density of $f(p_{iat} | s_{iat})$. The difference between the self-reported survival probability p_{iat} and the subjective survival probability S_{iat} is the measurement error.

The primary objective of this paper is to use the observed p_{iat} to update the prior density $\pi(s_{iat})$ and to obtain the posterior density $\pi(s_{iat} | p_{iat})$. After we observe p_{iat} , the posterior density of S_{iat} is given by:

$$\pi(s_{iat} | p_{iat}) = \frac{f(p_{iat} | s_{iat}) \pi(s_{iat})}{\int f(p_{iat} | s_{iat}) \pi(s_{iat}) ds_{iat}}.$$

If the loss function is given by: $L(S_{it}, \hat{S}_{it}) = E(S_{it} - \hat{S}_{it})^2$, the best estimator for S_{it} is $\hat{S}_{it} = E(S_{it} | p_{iat})$. We apply \hat{S}_{it} to the observed death record to obtain the model's parameter values. The log-likelihood function is given by:

$$\ln L = \sum_{\text{alive}} \ln \hat{S}_{it} + \sum_{\text{dead}} \ln(1 - \hat{S}_{it}). \quad (4)$$

We can obtain estimates of parameters of the prior and posterior densities by maximizing the log-likelihood function in (4).

IV. Estimation and Out-of-Sample Prediction

Maximizing the likelihood function in (4) requires specifying the distribution functions. The probability of an agent whose current age is a and who survives until age $a + t$ is given by s_{iat} . Different agents will have different survival probabilities. For the population of agents who share the same age a , their surviving probabilities are a random variable S_{iat} . Let the prior distribution for the random variable S_{iat} , $\pi(s_{iat})$ be the truncated normal distribution. We also let the mean of S_{iat} be $\exp(-\psi \Delta \Lambda_{0at})$, variance σ_2^2 , while the truncation range is $0 < s_{iat} < 1$. The prior distribution is given by:

$$\pi(s_{ia}; \psi) = \frac{\frac{1}{\sigma_2} \phi\left(\frac{s_{ia} - v_{ia}}{\sigma_2}\right)}{\Phi\left(\frac{1 - v_{ia}}{\sigma_2}\right) - \Phi\left(-\frac{v_{ia}}{\sigma_2}\right)}, \quad (5)$$

where v_{ia} and σ_2 satisfy the equation:

$$\exp(-\psi \Delta \Lambda_{0at}) = v_{iat} - \sigma_2 \eta(0, 1, v_{iat}, \sigma_2). \quad (6)$$

The right hand side of (6) represents the mean of the truncated normal in (5) with its formula provided in the Appendix. We let the conditional density of the responses to interview survival questions follow a censored normal distribution:

$$f(p_{ia\tau} | s_{ia\tau}) = \phi\left(\frac{p_{ia\tau} - \mu_{ia\tau}}{\sigma_1}\right) \text{ when } 0 < p_{ia\tau} < 1;$$

$$\Pr(p_{ia\tau} = 0 | s_{ia\tau}) = 1 - \Phi\left(\frac{\mu_{ia\tau}}{\sigma_1}\right) \text{ and } \Pr(p_{ia\tau} = 1 | s_{ia\tau}) = 1 - \Phi\left(\frac{1 - \mu_{ia\tau}}{\sigma_1}\right).$$

Furthermore, we assume that the expected value of the conditional distribution is s_{ia} . Thus, μ_{ia} and σ_1 satisfy the following equation:

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

The formula for the mean of the censored normal is given in the Appendix. The censored normal captures the idea that many observations may be at zero or one.

Given p_{iat} , the posterior distribution is given by:

$$\pi(s_{ia} | p_{ia\tau}) = \frac{\phi\left(\frac{p_{ia\tau} - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right)}{\int \phi\left(\frac{p_{ia\tau} - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}. \quad (7)$$

The distribution in (7) is no longer a normal or a censored normal. The best estimator for S_{ia} under a mean square loss function is its mean:

$$\hat{S}_{ia} = \int_0^1 s_{ia} \pi(s_{ia} | p_{ia\tau}) ds_{ia} = \frac{\int_0^1 s_{ia} \phi\left(\frac{p_{ia\tau} - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}{\int \phi\left(\frac{p_{ia\tau} - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}. \quad (8)$$

When $p_{iat} = 0$, we have

$$\begin{aligned} \Pr(s_{ia} < s | p_{ia\tau} = 0) &= \frac{\int_0^s \Pr(p_{ia\tau} = 0 | s_{ia}) \pi(s_{ia}; \psi) ds_{ia}}{\int_0^1 \Pr(p_{ia\tau} = 0 | s_{ia}) \pi(s_{ia}; \psi) ds_{ia}} \\ &= \frac{\int_0^s \left(1 - \Phi\left(\frac{\mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}{\int_0^1 \left(1 - \Phi\left(\frac{\mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}} \end{aligned}$$

Thus, the posterior distribution S_{ia} given $p_{ia\tau} = 0$ is:

$$\pi(s_{ia} | p_{ia\tau} = 0) = \frac{\left(1 - \Phi\left(\frac{\mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right)}{\int_0^1 \left(1 - \Phi\left(\frac{\mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}.$$

Then, the best predictor for S_{ia} when $p_{iat} = 0$ is:

$$\hat{S}_{ia} = \frac{\int_0^1 s_{ia} \left(1 - \Phi\left(\frac{\mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}{\int_0^1 \left(1 - \Phi\left(\frac{\mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}. \quad (9)$$

Similarly, when $p_{iat} = 1$,

$$\pi(s_{ia} | p_{ia\tau} = 1) = \frac{\left(1 - \Phi\left(\frac{1 - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right)}{\int_0^1 \left(1 - \Phi\left(\frac{1 - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}},$$

with the best predictor being given by:

$$\hat{S}_{ia} = \frac{\int_0^1 s_{ia} \left(1 - \Phi\left(\frac{1 - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}{\int_0^1 \left(1 - \Phi\left(\frac{1 - \mu_{ia}(s_{ia}, \sigma_1)}{\sigma_1}\right)\right) \phi\left(\frac{s_{ia} - v_{ia}(\psi, \sigma_2)}{\sigma_2}\right) ds_{ia}}. \quad (10)$$

In (8), (9) and (10), we obtain the predicted \hat{S}_{ia} given the observed subjective survival probability of $p_{ia\tau}$.

Since respondents are interviewed every two years, we update information regarding whether they are still alive accordingly. The likelihood function in (4) should be changed to: $\ln L = \sum_{\text{alive}} \ln \hat{S}_{ia2} + \sum_{\text{dead}} \ln(1 - \hat{S}_{ia2})$. However, the self-reported survival probability is not merely the survival probability during a two-year period. Rather, it typically represents a survival probability 10-15 years in the future.

From (1) and (2), γ_i in the hazard-scaling model can be calculated by:

$$\hat{\gamma}_i = -\frac{\ln(\hat{S}_{ia\tau})}{\Delta\Lambda_{0a\tau}}. \quad (11)$$

Therefore,

$$\hat{S}_{ia2} = \hat{S}_{ia\tau}^{\Delta\Lambda_{0a2}/\Delta\Lambda_{0a\tau}}, \quad (12)$$

After substituting the above expression of $\hat{\gamma}_i$, the log-likelihood function can be written as:

$$\ln L = \sum_{\text{alive}} \ln \hat{S}_{ia\tau}^{\Delta\Lambda_{0a2}/\Delta\Lambda_{0a\tau}} + \sum_{\text{dead}} \ln \left(1 - \hat{S}_{ia\tau}^{\Delta\Lambda_{0a2}/\Delta\Lambda_{0a\tau}} \right).$$

Although we cannot arrive at the explicit expression of γ_i through the age-scaling model, we can numerically solve the following equation to obtain the estimated value of γ_i .

$$\hat{S}_{ia\tau} = \exp \left[-\Lambda_0 \left(a + \frac{\tau}{\hat{\gamma}_i} \right) + \Lambda_0(a) \right].$$

Then,

$$\hat{S}_{ia2} = \exp \left[\Lambda_0 \left(a + \frac{\tau}{\hat{\gamma}_i} \right) - \Lambda_0 \left(a + \frac{2}{\hat{\gamma}_i} \right) \right] \hat{S}_{ia\tau}. \quad (13)$$

In sum, we let the prior survival probability distribution from age a to age $a+t$ be a truncated normal (between 0.0 and 1.0). Moreover, the conditional density of observed survival probabilities is assumed to be a censored normal, when wallowing for the focal points 0.0 and 1.0, while the posterior density of the survival probabilities will have a distribution that does not allow for the focal points 0.0 and 1.0. In order to obtain the model's parameter values, we apply the posterior distribution mean to actual death records between wave 1 and wave 2 in order to estimate a person's survival probability.

Both the hazard-scaling model in (2) and the age-scaling model in (3) are estimated. In each model, we first let $\psi = 1$, constraining the mean of the prior density to be the same as that of the life table. Whenever we do this, we refer to the model as the constrained model. In addition, we let ψ be a parameter to be estimated. In this case, we let the data determine if the prior density mean is the same as the life table survival probability. We refer to such a model as an unconstrained model. Table 3 lists the results of four different specifications: constrained hazard-scaling model, unconstrained hazard-

scaling model, constrained age-scaling model, and finally, unconstrained age-scaling model. All four specifications yield reasonable estimates that are highly significant. Moreover, likelihood ratio tests favor unconstrained models over constrained models.

Since we use a survey that has three-waves of data, we can apply the estimated parameters to actual survival experiences from wave 3 observations and compare the log-likelihood of each model in order to select the best one. We select the sample that are comprised of individuals who are still alive in wave 2, then calculate the log-likelihood values separated by those who are alive in wave 3 and those who are dead between waves 2 and 3. The log-likelihood from the out-of-sample prediction is given by:

$$\ln L = \sum_{\text{alive in wave 3}} \ln \hat{S}_{ia4} + \sum_{\text{dead b/w wave 2 \& 3}} \ln(1 - \hat{S}_{ia4})$$

The log-likelihood values from the out-of-sample predictions are reported in Table 3. The two unconstrained models perform much better than the corresponding two constrained models. Between the two unconstrained models, the hazard-scaling model yields slightly better likelihood values than the unconstrained age-scaling model.

For each specification, we calculate the optimistic indices γ_i for each individual. The formula to calculate γ_i in the hazard-scaling model is given by Equation (11) while the formula to calculate γ_i in the age-scaling model is provided in Equation (13). Table 4 presents the summary statistics of the indices and the correlation coefficients from the four different models' indices.

From Table 4, we find that the correlation coefficients among different indices are very high. The lowest correlation coefficient between the unconstrained hazard-scaling model and the age-scaling model is -.8975. The negative signs for the correlation coefficients between the two models are expected; that is, in the hazard-scaling model,

the larger the index, the less optimistic a person is. The inverse result can be found for the age-scaling model; namely, the larger the index, the more optimistic a person is. The highest correlation coefficient between the unconstrained and the constrained hazard scaling models is .9887, which is very close to 1.

Although the correlations among the four different specifications are very high, the means of estimated γ_i from the four different models differ significantly. These means are also reported in Table 4. The estimated $\hat{\gamma}_i$ for unconstrained specifications portrays a more optimistic picture than those for constrained specifications. In the hazard-scaling model, the average $\hat{\gamma}_i$ in the constrained specification is 1.020, indicating that an individual's subjective survival probability on average is very close to the life table. In the unconstrained version of the hazard model, the average $\hat{\gamma}_i$ is .822, indicating that people are generally optimistic about their survival probabilities. Similar patterns occur in the two specifications of the age-scaling model.

In the constrained specification, the means of the prior densities (truncated normal) are constrained according to life-table survival probabilities. The Bayesian update model only changes its σ_2 , i.e., the standard deviation of the original normal density that generates the truncated normal density (See (5) and (6)). Although updating σ_2 may have some effects on the means of the prior densities, the effect is minimal. Therefore it is not surprising to see that the constrained versions of both models are very similar to life tables. In the unconstrained specification, and in addition to obtaining the value of σ_2 , the updated Bayesian model also changes the mean of the prior densities through ψ .

Although different specifications yield different levels of optimistic indices, an important feature of all these indices is that a significant heterogeneity exists among all individuals. The individual heterogeneity in γ_i can be summarized by a simple regression that uses the optimism indices to regress certain demographic variables. In this regression, four different optimism indices represent dependent variables, while independent variables include a constant, the person's age, a male dummy, an African-American dummy, the Hispanic dummy, and lastly, a marriage status dummy. We report the estimation results in Table 5.

From the estimates reported in Table 5, the coefficients for African American dummies are negative for the hazard scaling model specifications and positive for the age-scaling model specifications. All coefficients indicate that African Americans are more optimistic than White respondents. No difference exists between Hispanic and White respondents in terms of their optimism indices. Another pattern that can be found in all four specifications is the finding that male respondents are more optimistic than female respondents. In addition, older respondents are generally more optimistic than younger respondents in three specifications. The only exception is the unconstrained age-scaling model where the age coefficient is insignificant. Finally, whether a person is married does not make any difference in his/her optimism indices.

Tables 6 through 9 provide the predicted survival probabilities of four different specifications, the stated survival probabilities, and the life table survival probabilities. The predicted survival probabilities in the unconstrained specifications are higher than those based on constrained specifications. This derives from the fact that the

unconstrained specifications produce indices that indicate more optimism than those based on unconstrained specifications.

In Figure 1, we produce two fitted probability histograms for males and females between the ages of 70-74 at the time the survey is conducted for the constrained hazard-scaling model. The histograms for all other age groups and all other models are the same save for their location. From this figure, all focal responses of zero and one have moved away from zero and one. For example, for males who are between 70 – 74 years old at the time of the survey, the predicted probabilities of surviving to age 85 are .22 and .61 if the responses are zero and one, respectively. Figure 2 has various survival curves for both males and females at age 70 for both constrained and unconstrained specifications in the hazard-scaling model. Graphs based on other models at other age categories look similar. In Figure 2, the lines *personal-p=1* and *personal-p=0* represent the survival curves if the response is one and zero, respectively. The line *personal-p=Average* represents the survival curve if the response represents the average of all responses. Not surprising, a person whose response is one typically has the highest survival curve, thus demonstrating the highest survival probabilities, while a person whose response is 0.0 has the lowest survival curve.

The densities of prior and posterior densities are illustrated in Figure 3. The first panel in Figure 3 shows the prior and posterior densities if the response is one, with the posterior density lying to the right of the prior density. Similarly, in the second panel in Figure 3, the posterior density lies to the left of the prior density if the response is zero. This is what one would expect from the Bayesian update model. The third panel in Figure 3 illustrates a case where the response is 0.5. Finally, we produce histograms of the

estimated optimistic parameters γ_i for all four models in Figure 4. The average and standard deviations of γ_i are also given in the histograms. There are significant variations between these indices. The significant variations in optimistic indices produce significant variations in individual survival curves.

IV. Conclusions

Many economic models are based on forward-looking behavior on the part of economic agents. Surveys such as HRS and AHEAD ask individuals for their expectations on the probability of given future events occurring in their lifetime. On average, the subjective probability of a future event is consistent with the observed probability that the event does occur. In other words, individual survival probabilities are consistent with those from the life tables.

However, at the micro level, the subjective probability responses in HRS and AHEAD suffer serious problems with focal responses of 0.0 and 1.0. Consequently, applying subjective probabilities are extremely limited if “true” subjective survival probabilities are not recovered.

In this paper, we suggest a Bayesian update model that is based on actual survival information to account for problems caused by focal responses of 0.0 and 1.0. As a result, individual survival curves derived from the model do not suffer the problems of focal responses. We also propose two approaches to model the individual heterogeneities of their subjective survival curves. One approach modifies the life table hazard rates while another approach models the subjective aging process that is different from the life table aging process. The model is estimated from the observed survival information of our sample. From the estimated model, we construct several optimistic indices for each

individual and conduct a test that is based on out-of-sample prediction. These optimistic are used to create individual subjective survival curves that have considerable variations and are readily applicable to economic models that require individual subjective survival curves. In a companion paper, we apply these individual subjective survival curves to a life-cycle model of savings and bequests.

Appendix

Mean of the Truncated Normal Distribution.

If $x \sim N[\mu, \sigma^2]$ and e and f are constant,

$$E[x|e \leq x \leq f] = \mu - \sigma \eta(e, f, \mu, \sigma),$$

$$\text{where } \eta(e, f, \mu, \sigma) = \frac{\phi\left(\frac{f-\mu}{\sigma}\right) - \phi\left(\frac{e-\mu}{\sigma}\right)}{\Phi\left(\frac{f-\mu}{\sigma}\right) - \Phi\left(\frac{e-\mu}{\sigma}\right)}.$$

Proof: $x \sim N[\mu, \sigma^2]$, the density of the truncated normal distribution is then

$$g(x|e \leq x \leq f) = \frac{\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{f-\mu}{\sigma}\right) - \Phi\left(\frac{e-\mu}{\sigma}\right)}.$$

The mean is

$$\begin{aligned}
E[x|e \leq x \leq f] &= \int_e^f x g(x|e \leq x \leq f) dx = \frac{\int_e^f x \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx}{\Phi\left(\frac{f-\mu}{\sigma}\right) - \Phi\left(\frac{e-\mu}{\sigma}\right)} \\
&= \mu - \sigma \frac{\phi\left(\frac{f-\mu}{\sigma}\right) - \phi\left(\frac{e-\mu}{\sigma}\right)}{\Phi\left(\frac{f-\mu}{\sigma}\right) - \Phi\left(\frac{e-\mu}{\sigma}\right)} \\
&= \mu - \sigma \eta(e, f, \mu, \sigma)
\end{aligned}$$

Mean of the Censored Normal Distribution.

If $x^* \sim N[\mu, \sigma^2]$ and $x = e$ if $x^* \leq e$, or $x = x^*$ if $e \leq x^* \leq f$, or $x = f$ if $x^* \geq f$, where e and f are constant, then

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

Proof: The conditional mean,

$$\begin{aligned}
E[x] &= \Pr(x = e) \times E[x|x = e] + \Pr(e < x < f) \times E[x|e < x < f] + \Pr[x \geq f] \times E[x|x = f] \\
&= \Pr(x^* \leq e) \times e + \Pr(e < x^* < f) \times E[x^*|e < x^* < f] + \Pr[x^* \geq f] \times f \\
&= \Phi\left(\frac{e-\mu}{\sigma}\right)e + \left[\Phi\left(\frac{f-\mu}{\sigma}\right) - \Phi\left(\frac{e-\mu}{\sigma}\right)\right][\mu - \sigma\eta(e, f, \mu, \sigma)] + \left[1 - \Phi\left(\frac{f-\mu}{\sigma}\right)\right]f
\end{aligned}$$

using the properties of the truncated normal mean.

References

Bassett, William. and Robin Lumsdaine. "Probability Limits – Are Subjective Assessments Adequately Accurate?" Journal of Human Resources, Vol 36 (2) (Spring 2001): 327-363.

Hamermesh, Daniel. "Expectations, Life Expectancy, and Economic Behavior," Quarterly Journal of Economics, Vol 10, No 2 (May 1985): 389-408.

- Hubbard, R. Glenn, Jonathan Skinner and Stephen Zeldes. "Precautionary Saving and Social Insurance," Journal of Political Economy, Vol 103, No 2 (April 1995): 360-99.
- Hurd, Michael D. "Mortality Risks and Bequests," Econometrica, Vol 57, n4 (July 1989): 779-813.
- Hurd, Michael D, Daniel McFadden and Li Gan. "Subjective Survival Curves and Life Cycle Behavior." In David Wise, ed, Inquiries of Economics of Aging, Chicago: The University of Chicago Press (1998), 259-305.
- Hurd, Michael D and Kathleen McGarry. "Evaluation of Subjective Probability Distributions in the Health and Retirement Study." Journal of Human Resources, Vol 30 (Suppl. 1995).
- Manski, Charles. "Dynamic Choice in Social Settings: Learnings from the Experiences of Others." Journal of Econometrics, Vol 58, No1-2 (July 1993): 121-36.

Table 1 Self-Reported and Life Table Survival Probabilities

	Target Age									
	Male					female				
	80	85	90	95	100	80	85	90	95	100
	<u>Wave 1</u>									
Means										
AHEAD	0.557	0.510	0.382	0.332	0.302	0.570	0.510	0.386	0.307	0.289
Life Table	0.593	0.422	0.252	0.114	0.037	0.716	0.605	0.432	0.232	0.081
Median										
AHEAD	0.500	0.500	0.400	0.250	0.100	0.500	0.500	0.450	0.100	0.100
Life Table	0.593	0.422	0.252	0.115	0.037	0.723	0.603	0.433	0.232	0.076
# of obs	90	951	631	436	175	575	1334	978	664	309
	<u>Wave 2</u>									
Means										
AHEAD	0.524	0.279	0.622	0.278	0.574	0.516	0.283	0.692	0.296	0.559
Life Table	0.614	0.457	0.284	0.138	0.078	0.736	0.633	0.464	0.260	0.138
Median										
AHEAD	0.500	0.200	0.600	0.200	0.600	0.500	0.200	0.600	0.200	0.600
Life Table	0.629	0.456	0.285	0.140	0.051	0.746	0.632	0.465	0.261	0.100
# of obs	95	1044	675	451	223	620	1436	1090	807	498

Table 2 Focal Responses

		Wave 2				Total
		continuous	0	1	NA	
Wave 1	continuous	2,700	328	221	153	3,413
		79.3%	9.6%	6.5%	4.6%	41.5%
	0	1,599	201	177	104	2,041
		76.4%	9.9%	8.7%	5.1%	24.8%
	1	452	60	64	29	605
		74.7%	9.9%	10.6%	4.8%	7.4%
	NA	1,467	169	350	173	2,159
		68.0%	7.8%	16.2%	8.0%	26.3%
	Total	6,189	758	812	459	8,218
		75.3%	9.2%	9.9%	5.6%	

Table 3: Estimation Results

Parameter	Hazard-scaling		Age-scaling	
	$\psi=1$	ψ is a parameter	$\psi=1$	ψ is a parameter
Std dev of conditional density: σ_1 (censored normal)	.3255 (.1197) ^a	.1837 (.0154)	.5434 (.0012)	.2793 (.0312)
Std dev of prior density: σ_2 (truncated normal)	.2045 (.0045)	.1165 (.0176)	.3159 (.0000011)	0.1083 (.0304)
Average optimistic parameter: ψ		.7226 (.0507)		0.6590 (.0011)
Maximum Likelihood Value:	-1495	-1483	-1500	-1491
Log likelihood for out-of-sample Prediction	-1692.9	-1532.4	-1644.1	-1533.4

a) Standard errors are in parenthesis.

Table 4: Correlation Coefficients among Four Optimistic Indices

	Hazard-scaling (constrained) ^a	Hazard-scaling (unconstrained)	Age-scaling (constrained)	Age-scaling (unconstrained)
Hazard-scaling (constrained)	1	.9887 (.00024) ^b	-.9000 (.0014)	-.9019 (.0018)
Hazard-scaling (unconstrained)		1	-.8975 (.0017)	-.9284 (.0016)
Age-scaling (constrained)			1	.9479 (.0015)
Age-scaling (unconstrained)				1
Means (std error)	1.040 (.375)	.822 (.296)	1.051 (.227)	1.271 (.186)

a) “constrained” means $\psi=1$; “unconstrained” means ψ is a parameter to be estimated.

b) Standard errors are in parenthesis, calculated from bootstrapping 1,000 times of the sample.

Table 5: Summary Regressions of Four Optimistic Indices

	Hazard-scaling (<i>constrained</i>)	Hazard-scaling (<i>unconstrained</i>)	Age-scaling (<i>constrained</i>)	Age-scaling (<i>unconstrained</i>)
Constant	1.120 (.38) ^{a)}	1.546 (.050)	.740 (.039)	1.201 (.189)
Age	-.011 (.00081)	-.0089 (.00064)	.0036 (.00050)	-.0036 (.0051)
Male	-.176 (.010)	-.132 (.0079)	.0982 (.0059)	.0914 (.00475)
Black	-.065 (.022)	-.052 (.012)	.0625 (.0091)	.0446 (.0074)
Hispanic	.065 (.022)	.0043 (.017)	.0021 (.014)	-.00467 (.011)
Married	.0066 (.0104)	.0072 (.0082)	-.0095 (.0063)	.0034 (.0052)
No. of obs.	6089	6089	6089	6089
R ²	.0948	.0909	.0638	.0859

a) Standard errors are in parenthesis.

Table 6: Fitted Survival Probabilities
(Constrained hazard-scaling model)

Age Group	Target Age	Life Table	Nonfocal Respondents		Focal Respondents		All Respondents	
			Predicted	Stated	$p_{iar} = 0$ Predicted	$p_{iar} = 1$ Predicted	Predicted	Stated
Female								
70-74	85	0.5880	0.5565 (0.0696) ^{a)}	0.5001	0.3571 (0.0218)	0.7592 (0.0213)	0.5604 (0.1215)	0.5095
75-79	90	0.4250	0.4426 (0.0745)	0.4616	0.2486 (0.0155)	0.6584 (0.0348)	0.4107 (0.1319)	0.3885
80-84	95	0.2240	0.2904 (0.0666)	0.4139	0.1398 (0.0176)	0.4806 (0.0213)	0.2485 (0.1113)	0.3029
Male								
70-74	85	0.3970	0.4293 (0.0680)	0.4845	0.2270 (0.0225)	0.6342 (0.0250)	0.4383 (0.1199)	0.5103
75-79	90	0.2500	0.3086 (0.0651)	0.4127	0.1466 (0.0133)	0.5091 (0.0446)	0.2936 (0.1079)	0.3820
80-84	95	0.1130	0.1848 (0.0543)	0.3960	0.0771 (0.0119)	0.3208 (0.0561)	0.1645 (0.0845)	0.3324

a) Standard errors are in parenthesis.

Table 7: Fitted Survival Probabilities
(Unconstrained hazard-scaling model)

Age Group	Target Age	Life Table	Nonfocal Respondents		Focal Respondents	Focal Respondents	All Respondents	
			Predicted	Stated	$p_{iar} = 0$ Predicted	$p_{iar} = 1$ Predicted	Predicted	Stated
Female								
70-74	85	0.5880	0.6266 (0.0708) ^{a)}	0.5001	0.4659 (0.0193)	0.8062 (0.0207)	0.6322 (0.1084)	0.5095
75-79	90	0.4250	0.5171 (0.0769)	0.4616	0.3602 (0.0169)	0.7069 (0.0323)	0.4927 (0.1162)	0.3885
80-84	95	0.2240	0.3634 (0.0790)	0.4139	0.2117 (0.0242)	0.5551 (0.0158)	0.3213 (0.1167)	0.3029
Male								
70-74	85	0.3970	0.5042 (0.0710)	0.4845	0.3339 (0.0304)	0.6838 (0.0227)	0.5129 (0.1092)	0.5103
75-79	90	0.2500	0.3814 (0.0750)	0.4127	0.2224 (0.0199)	0.5777 (0.0377)	0.3667 (0.1109)	0.3820
80-84	95	0.1130	0.2673 (0.0771)	0.3960	0.1191 (0.0171)	0.4503 (0.0380)	0.2390 (0.1146)	0.3324

a) Standard errors are in parenthesis.

Table 8: Fitted Survival Probabilities
(Constrained age-scaling model)

Age Group	Target Age	Life Table	Nonfocal Respondents		Focal Respondents	Focal Respondents	All Respondents	
			Predicted	Stated	$p_{iar} = 0$ Predicted	$p_{iar} = 1$ Predicted	Predicted	Stated
Female								
70-74	85	0.5880	0.5561 (0.0565)	0.5001	0.5589 (0.1167)	0.7554 (0.0195)	0.5589 (0.1167)	0.5095
75-79	90	0.4250	0.4452 (0.0611)	0.4616	0.2516 (0.0141)	0.6629 (0.0342)	0.4136 (0.1276)	0.3885
80-84	95	0.2240	0.2917 (0.0524)	0.4139	0.1505 (0.0165)	0.4750 (0.0249)	0.2529 (0.1024)	0.3029
Male								
70-74	85	0.3970	0.4306 (0.0552)	0.4845	0.2320 (0.0209)	0.6399 (0.0250)	0.4407 (0.1154)	0.5103
75-79	90	0.2500	0.3112 (0.0524)	0.4127	0.1572 (0.0132)	0.5070 (0.0476)	0.2974 (0.0998)	0.3820
80-84	95	0.1130	0.1788 (0.0431)	0.3960	0.0850 (0.0123)	0.2879 (0.0646)	0.1602 (0.0717)	0.3324

a) Standard errors are in parenthesis.

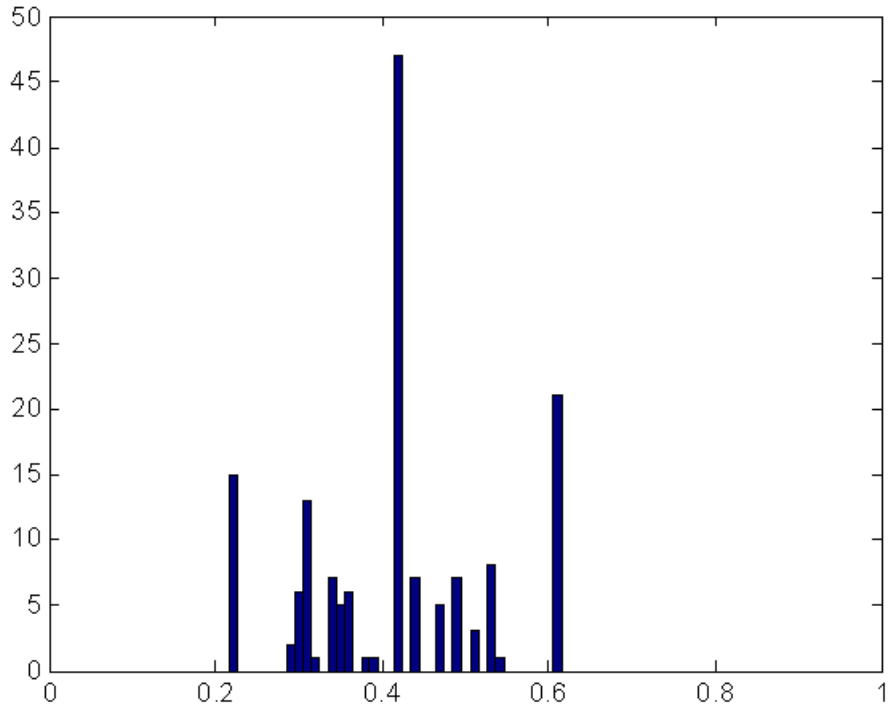
Table 9: Fitted Survival Probabilities
(Unconstrained age-scaling model)

Age Group	Target Age	Life Table	Nonfocal Respondents		Focal Respondents	Focal Respondents	All Respondents	
			Predicted	Stated	$p_{iar} = 0$ Predicted	$p_{iar} = 1$ Predicted	Predicted	Stated
Female								
70-74	85	0.5880	0.6689 (0.0384)	0.5001	0.5850 (0.0202)	0.7728 (0.0241)	0.6733 (0.0606)	0.5095
75-79	90	0.4250	0.5531 (0.0446)	0.4616	0.4720 (0.0184)	0.6592 (0.0369)	0.5413 (0.0649)	0.3885
80-84	95	0.2240	0.3825 (0.0457)	0.4139	0.3007 (0.0286)	0.4864 (0.0179)	0.3598 (0.0662)	0.3029
Male								
70-74	85	0.3970	0.5342 (0.0394)	0.4845	0.4427 (0.0351)	0.6329 (0.0258)	0.5392 (0.0613)	0.5103
75-79	90	0.2500	0.4043 (0.0458)	0.4127	0.2812 (0.0684)	0.5122 (0.0429)	0.3956 (0.0657)	0.3820
80-84	95	0.1130	0.2715 (0.0488)	0.3960	0.1738 (0.0239)	0.3719 (0.0417)	0.2509 (0.0728)	0.3324

a) Standard errors are in parenthesis.

Figure 1: Histograms of predicted survival probabilities

Predicted Survival Probabilities to age 85 among 70-74 year males: Constrained Hazard-Scaling Model



Predicted Survival Probabilities to age 85 among 70-74 year females: Constrained Hazard-Scaling Model

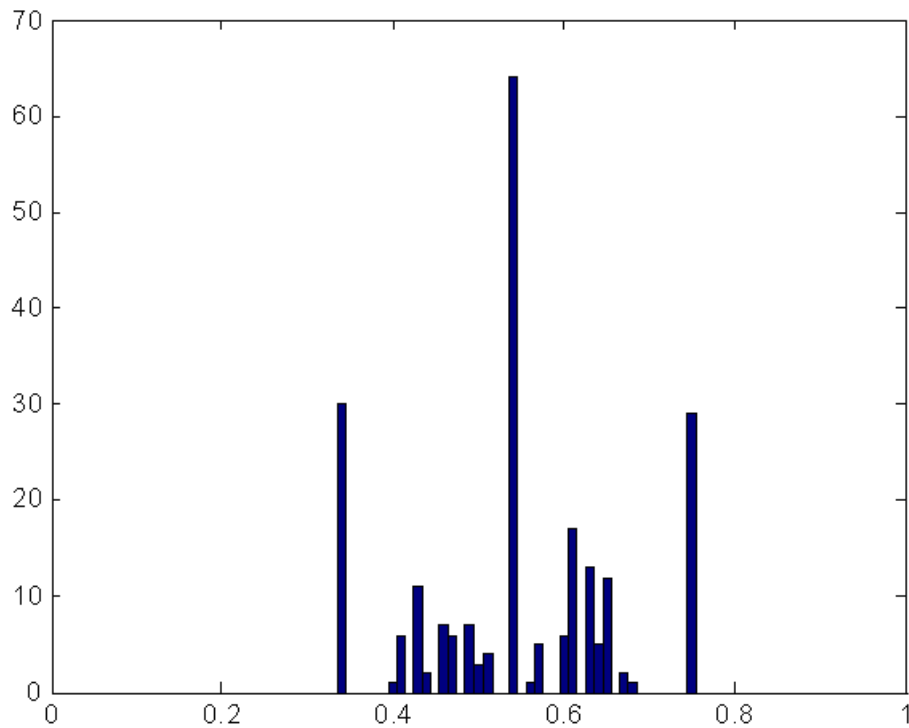
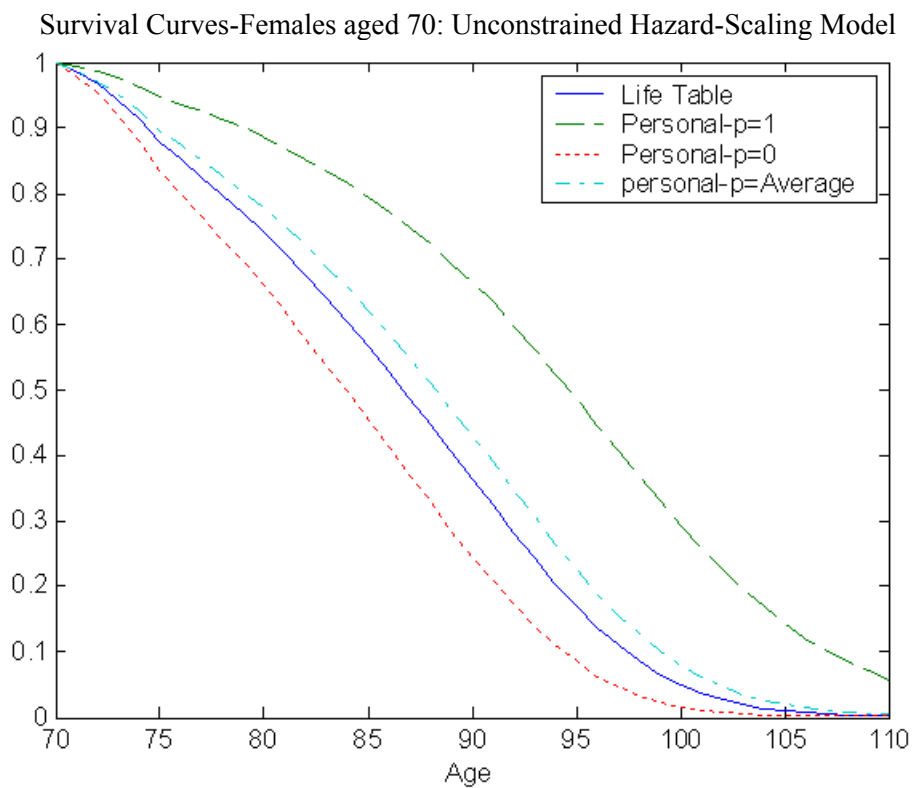
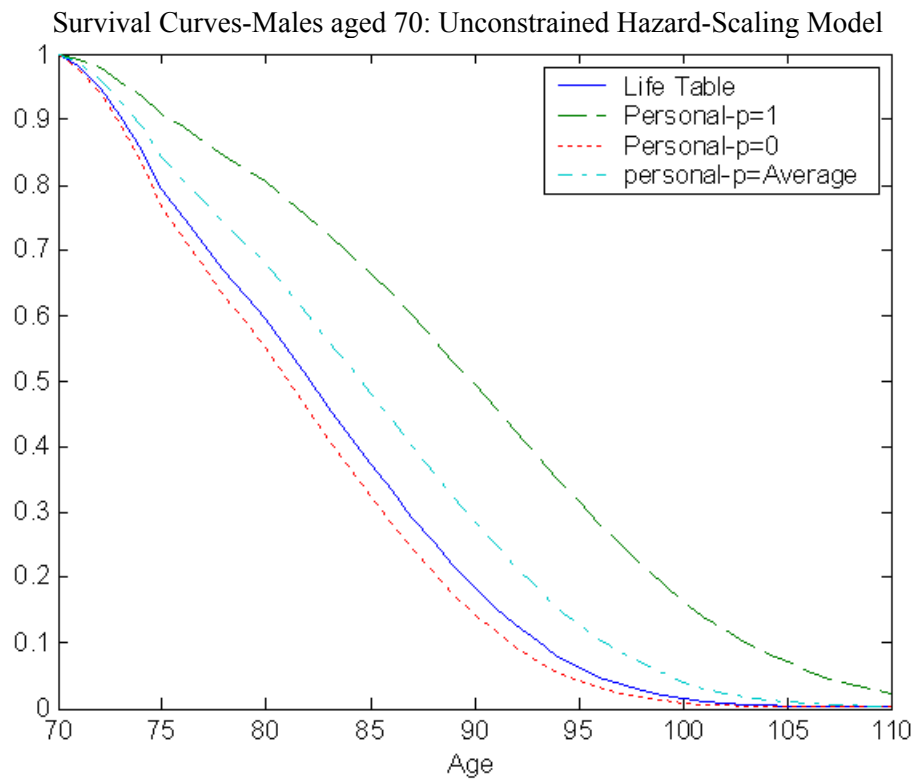
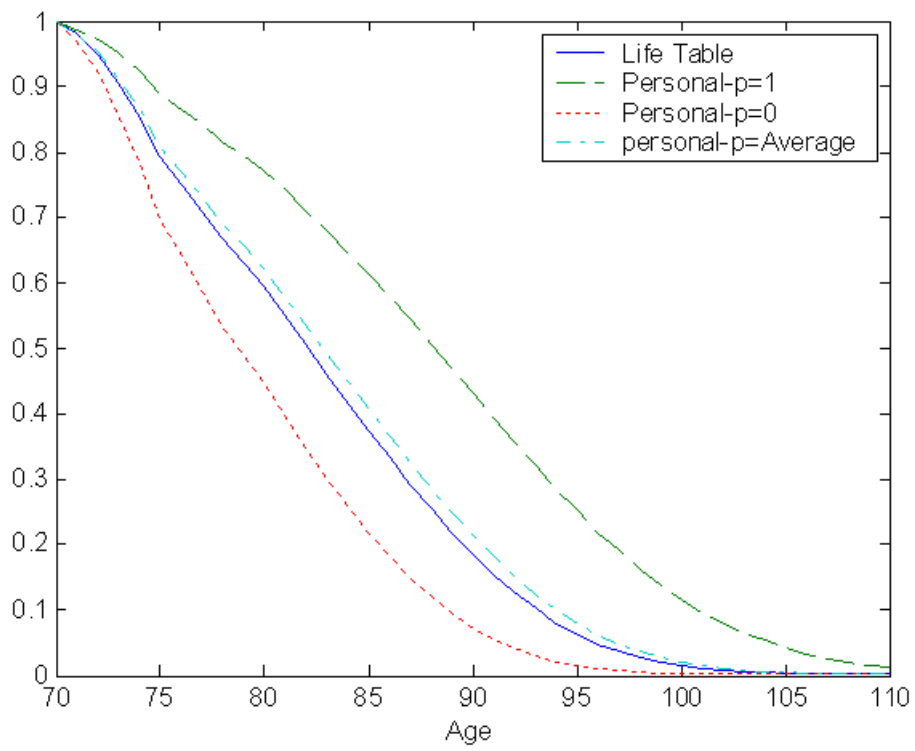


Figure 2: Survival curves



Survival Curves-Males aged 70: Constrained Hazard-Scaling Model



Survival Curves-Females aged 70: Constrained Hazard-Scaling Model

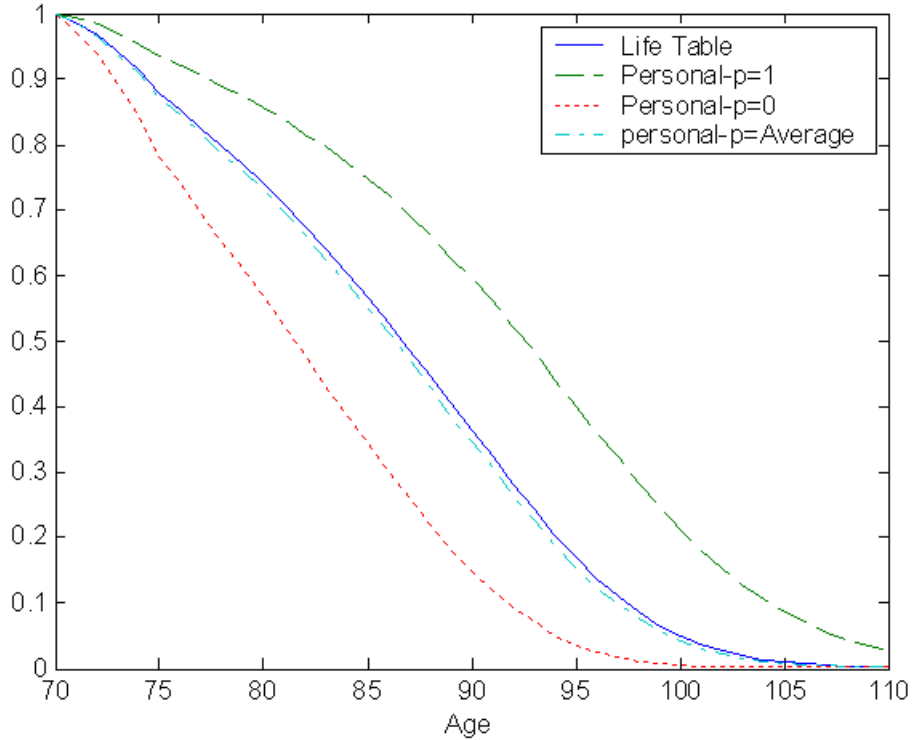
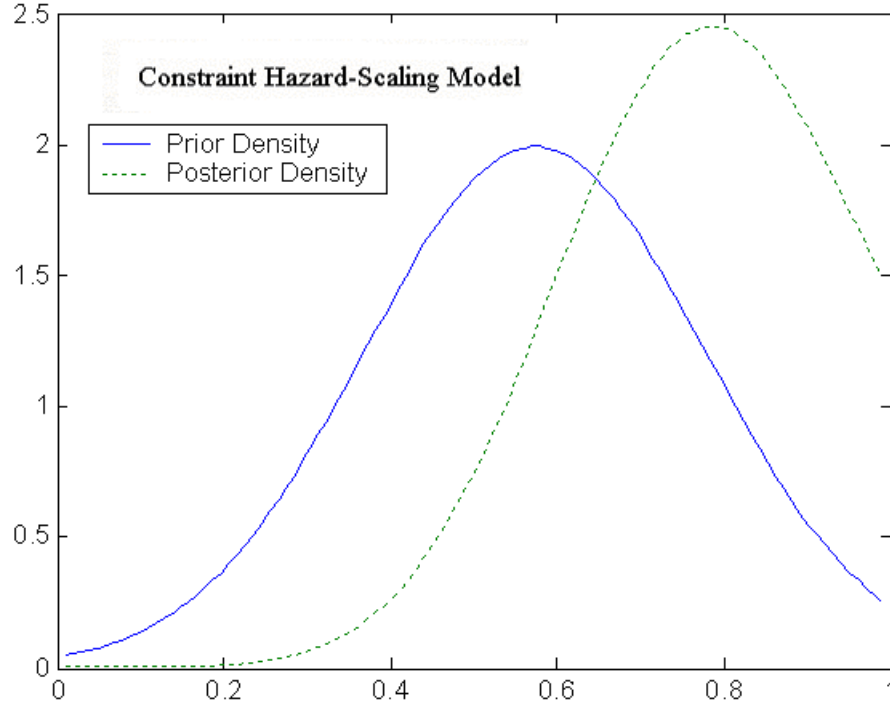
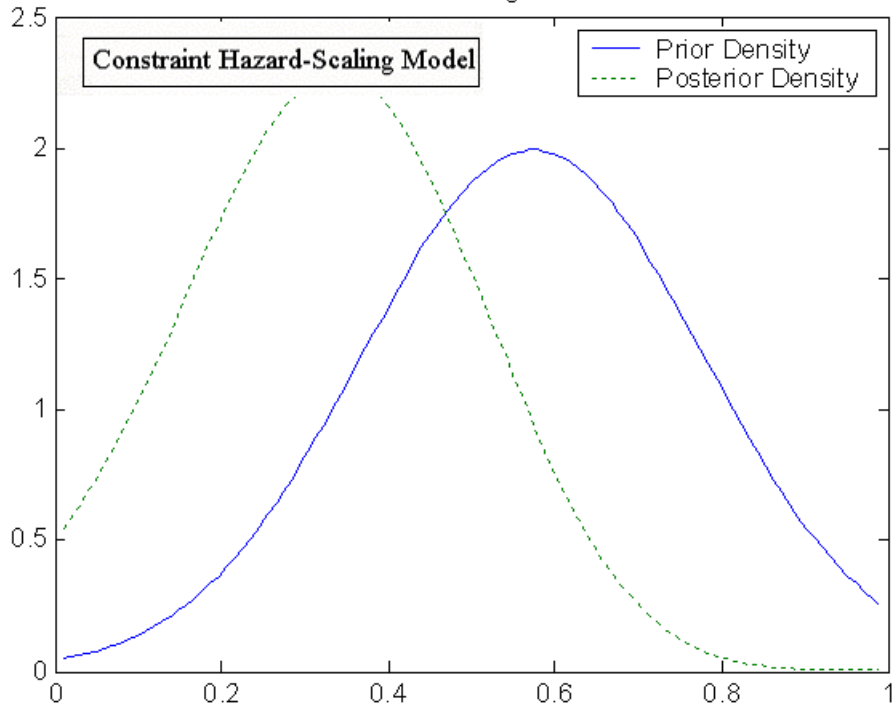


Figure 3: Densities of prior, conditional and posterior distributions

Prior and Posterior Densities for Males Aged 70 with $P = 1$: survival rate = 0.5654



Prior and Posterior distributions for Males Aged 70 with $P = 0$: survival rate = 0.5654



Prior and Posterior Densities for Males Aged 70 with P = 0.5: survival rate = 0.5654

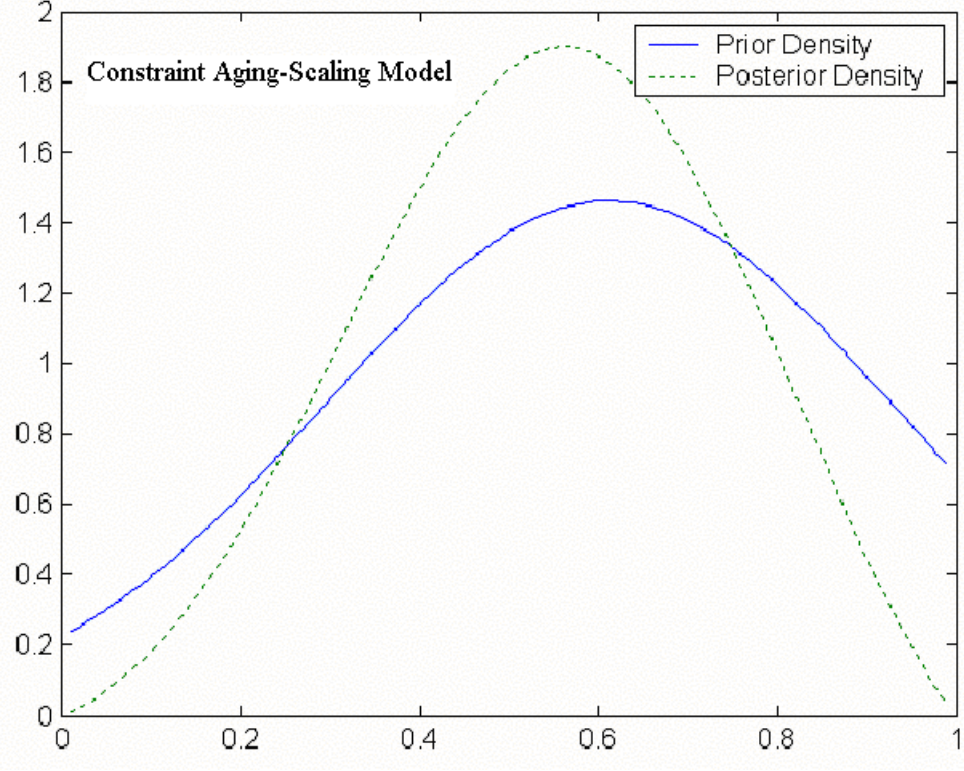
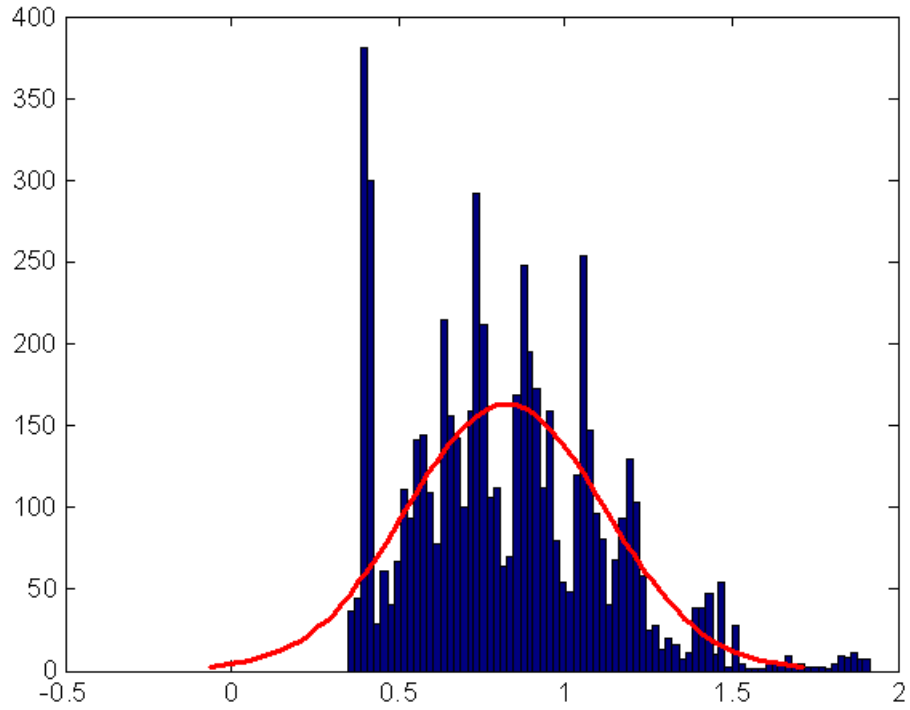
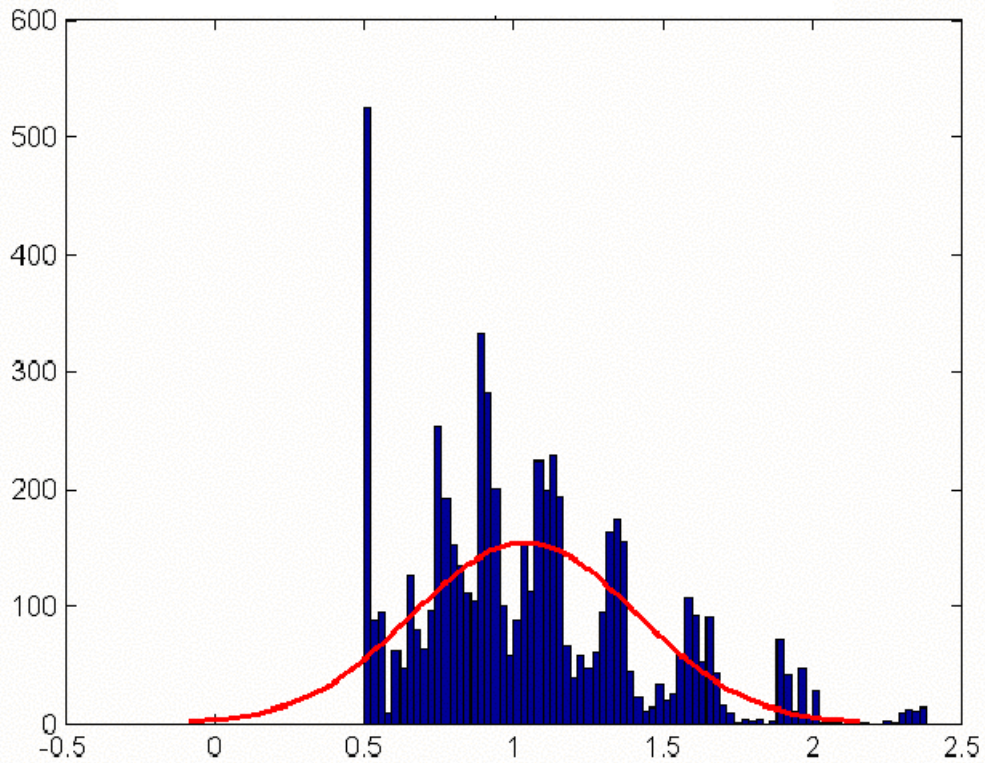


Figure 4: Histogram for optimistic indices

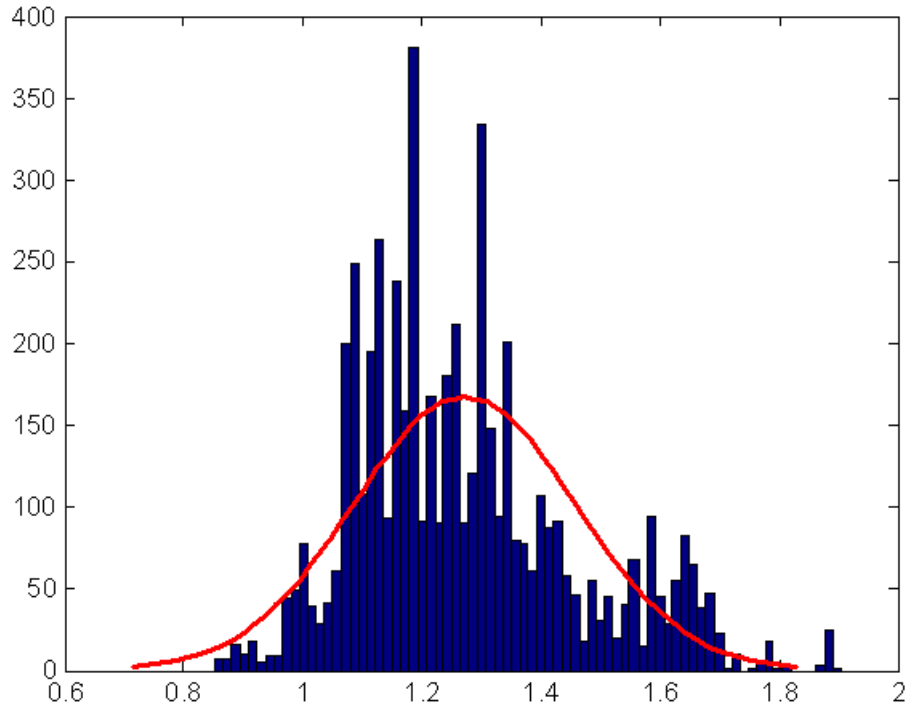
Histogram for optimistic index: unconstrained hazard-scaling model
mean = 0.8229, standard deviation = 0.2956



Histogram for optimistic index: constrained hazard-scaling model
mean = 1.0398, standard deviation = 0.3752



Histogram for optimistic index: unconstrained age-scaling model
mean = 1.2708, standard error = 0.1855



Histogram for optimistic index: constrained age-scaling model
mean = 1.0617, standard deviation = .2049

