Netspar

Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Ying Yang

Method for Calculating Healthy Life Expectancy by Including Dynamic Changes of Both Mortality and Health

Research Master Thesis 2010

Method for Calculating Healthy Life Expectancy by Including Dynamic Changes of Both Mortality and Health

Ying YANG*

August 4, 2010

Abstract

This thesis incorporates the self-reported health status information from the National Health Interview Survey in the United States into a cohort life table to estimate and forecast healthy life expectancy, which is the average of years lived in good health. First, the thesis defines the Health Status Index (HSI) representing the proportion of the population of people who are in bad health. Applying the HSI, the main contribution of this thesis is modeling the dynamic changes of both the mortality and health processes by using the Lee-Carter model and constructing their stochastic projections. Based on goodness-of-fit tests we find that the Lee-Carter model fits the data quite well. Healthy life expectancy (HLE) is estimated and projected using Sullivan's method by including the stochastic projectuon of the HSI into cohort life tables. The results show increasing trends of both life expectancy (LE) and healthy life expectancy (HLE), whereas the latter increases faster than the former. Another novelty of this thesis is the inclusion of uncertainty intervals by means of simulation method for expected simulated LE and HLE. We found that HLE have larger uncertainty than LE. Moreover, males's LE and HLE are lower than females' but increase faster with larger confidence intervals. The thesis also provides a comparison between models using level and logit HSI formats, and shows that healthy life expectancies derived from the models with logit HSI are slightly lower, and increase slower with narrower confidence intervals than from the level format models, and a logit transformation is superior to the level format by construction.

Keywords: Health Status Index, Mortality, Lee-Carter Model, Life Table, Life Expectancy, Healthy Life Expectancy, Uncertainty.

1 Introduction

In the past century, the elderly population of most of the highly developed countries, such as the United States, has increased steadily both in absolute terms and as a percentage of the total population, whereas mortality rates has declined dramatically. Such an aging trend brings significant effects for private and public pension programs, the social security fund, and the health care system. Costa (2002) found functional limitation of the U.S. people fell annually from the early twentieth century to the early 1990s. Similarly, Duggan and Imberman

*Research Master Thesis of Econometrics, Tilburg University. Contact: yangyingtina@gmail.com. Supervisors: Prof. Anja De Waegenaere and Prof. Bertrand Melenberg.

(2006) examined the trends in self-reported health provided by the National Health Interview Survey (NHIS) and found that health has improved on average for adults aged 50-64. In this context, when concerning the retirement policy and the health care system, society's attention no long purely stays on the increased life expectancy, but also on whether the increase is because of a growth in the number of healthy years, in the number of unhealthy years, or both. People's remaining lifetime lived in good health is usually called *healthy life expectancy* (HLE). Healthy life expectancy is also named *disability free life expectancy* by Sullivan (1971) or *active life expectancy* by Katz, Branch, Branson, Papsidero, Beck, and Greer (1983) and Manton, Corder, and Stallard (1993), as the period of life free of disability in activities of daily living. Healthy life expectancy is often used to measure the person-year without burden of functional disability both in the U.S. elderly population and for international comparison of developed countries with relative high life expectancies and aging population. This thesis is going to incorporate health information in the United States into a cohort life table to estimate and forecast healthy life expectancy. The thesis will first model the stochastic dynamic changes of both the mortality and health processes. Life expectancy is then estimated and forecasted from the cohort life table through a stochastic projection of mortality rates. Moreover, healthy life expectancy can be estimated and forecasted by combining cohort life tables and a stochastic projection of the health. For this purpose, for the remainder for this section, I will first start with the literature about how to measure the health, and the data set which researchers normally use. Then, the current research of Sullivan's model to estimate healthy life expectancy will be discussed. Finally, the models which the current researchers adopted for describing the health changes will be discussed, and the model used in this thesis to estimate and stochastically project the health process will be addressed, which is the novelty of this thesis.

The health status is not an easy-defined concept. Some researchers argue that it should be multidimensional and dynamic. To this extent, a nonparametric Grade of Membership (GoM) method is developed by Manton and Woodbury (1982), since it handles the multi dimensionality problem caused by many health information factors and also takes the variability of the degree of specific disorders into account. GoM fits the data better than did latent class models in the analysis of psychiatric diagnoses, both in general population samples, showed by Woodbury and Manton (1989), and in nursing home populations, showed by Manton, Cornelius, and Woodbury (1995). On the other hand, many researchers support the importance to examine the self-reported health status, which implies the perception of people themselves about their working abilities. Such as Lechner and Vazquez-Alvarez (2003) who used self-reported information on the assessed degree of disability for Germany addressed that becoming disabled reduces the probability of being in employment by around 9%. Lakdawalla, Goldman, and Bhattacharya (2004) analyzed the validation of the self-reported health condition to the ability to work. Gmez and Nicols (2006) examined how a self-reported health affects the probability of working for the Spanish population and found that there is a large probability that people quit the labor market when reporting bad health. In order to estimate healthy life expectancy which is more related to people's ability of working, the thesis will adopt the self-reported health information from the National Health Interview Survey (NHIS) in the United States to measure the health status.

To determine the health status for a specific cohort, many researchers use longitudinal data, for example, Manton, Stallard, and Corder (1997), Manton and Land (2000), and Manton, Gu, and Lowrimore (2008) used the National Long Term Care Surveys (NLTCS) longitudinal data, in which persons are longitudinally followed to the time of death. And, Portrait, Lindeboom, and Deeg (2001) modeled the health status and mortality jointly using the Dutch data from Longitudinal Aging Study Amsterdam (LASA), and employed a nonlinear panel data model in which health depends on the different grades of membership of health variables

and on a range of demographic and socioeconomic characteristics. However, the analysis based on longitudinal data is difficult to be duplicated in other countries, since those data are hard to obtain. On the other hand, much analysis on determinants of health status was performed mainly with cross-sectional health data due to the limited data availability. For example, Manton and Stallard (1991) combined health status and demographic and socioeconomic characteristics by cross section analysis and estimated healthy life expectancy for the U.S. elderly people using the GoM method. They first identified the various health dimensions using the GoM method and derived life expectancy for specific age-gender populations. Then they combined these two elements to derive healthy life expectancy.

After identifying the people's health status, we are able to estimate and forecast healthy life expectancy by combining health information with a state dependent life table. The most widely used method for healthy life expectancy is proposed by Sullivan (1971) by combing mortality information from a period life table and disability information from a cross-sectional disability survey, which is easy to obtain, to recalculate a period life table free of disability in the given age interval and compared it with the general method for calculating life expectancy. Sullivan's method allows to distinguish the expectation of residual lifetime free of disability and the expectation of disability, by introducing a disability weighting factor - average fraction of the year persons of that age group are free of disability. This method is commonly applied by researchers. For example, Manton, Gu, and Lamb (2006) presented estimates of changes in life expectancy and healthy life expectancy using Sullivan's method from 1935 to 1999 by including period-specific sequential cross-sectional disability prevalence data from the NLTCS and the NHIS into life tables. They suggested that Medicare and Medicaid benefits, which may have been partly responsible for the large recent increase in healthy life expectancy. Sullivan's method is the most appropriate and of great use to derive healthy life expectancy. Mathers and Robine (1997) and Livre, Brouard, and Heathcote (2003) used simulation method to test the performance of Sullivan's method and found that under stationarity assumptions, Sullivan's method, based on period life tables, provides consistent estimator of disability free life expectancy. Recently, Imai and Soneji (2007) built a statistical foundation of Sullivan's method and proved that Sullivan's method is unbiased and consistent without stationarity assumptions when using cohort life tables. For this reason, and due to the limited availability of longitudinal data set, the thesis employs Sullivan's method to estimate and forecast healthy life expectancy by combining the consecutive cross-sectional health information from the NHIS and cohort life tables. One difference with the original Sullivan's method is that instead of using disability data, I use the self-reported health status from the NHIS, and adopt the *health status index* (HSI) reflecting people's self-assessed bad health, which is more relevant to people's working ability. This measure refines the decomposition of life expectancy to the healthy and unhealthy part instead of the disabled and disability free part.

To estimate and forecast healthy life expectancy, it is necessary to model the dynamic changes of not only the mortality, but also and health processes. The current literature has already incorporated the health status into the life expectancy estimation. However, only a small part of the literature examines health changes in individuals over time. Manton, Stallard, and Tolley (1991) modeled the health by introducing multiple time-varying chronic disease risk factors and including it into the life expectancy analysis. This method shows that the health process and mortality jointly affect people's remaining lifetime. Portrait, Lindeboom, and Deeg (2001) adopted the panel data analysis to model the changes of the health status by a limited set of interpretable variables. Their model allows correlations between mortality and health status by unobserved individual factors. As a consequence, they are able to calculate the expected residual lifetime in a specific health status. The approach to model the health used in this thesis differs from theirs: this thesis undertakes the stochastic methodology proposed by Lee and Carter (1992) to model the dynamics of the health process directly. Moreover, the

thesis stochastically projects the health status through the Lee-Carter model and combines the stochastic projection with cohort life tables to estimate and forecast healthy life expectancy. The Lee-Carter model is a stochastic approach normally used to describe mortality changes and its future trend, see Tuljapurkar, Li, and Boe (2000), Lee and Miller (2001), Renshaw and Haberman (2003b), Renshaw and Haberman (2005), and many others. Applying the Lee-Carter approach to model the health process and deriving its stochastic projection to estimate and forecast healthy life expectancy is a contribution of this thesis to the current literature.

The next section will explain in detail estimating life expectancy and healthy life expectancy. The measure based on a period life table, originally proposed by Sullivan (1971) will be first illustrated, then a cohort life table with a time component will be addressed as the method used in this study. The health status index will be explained in this section as well. Section three introduces the Lee-Carter model and its estimation, first in the mortality context. Then how to model and project the health status index using the Lee-Carter approach is illustrated later in this section. Data used and the empirical analysis on mortality and health of the U.S. using the Lee-Carter model are described in section four, in which life expectancy and healthy life expectancy are estimated and projected based on stochastic projections of mortality and health. Furthermore, process risk and parameter risk are examined in the forecasting analysis. The last section concludes and outlines research questions for the future.

2 Life Expectancy and Healthy Life Expectancy

The average number of years of life remaining at a certain age of an individual is called *life expectancy* (LE). Life expectancy has shown an impressive rise during the last century in the United States. For example, the U.S. National Vital Statistics Report published that the expected remaining lifetime at birth for the total population using a period life table, increases from 49.24 years in 1900 to 65.47 years by 1950, and to 74.9 in the second half of the century. However, the continuing increase in life expectancy causes a rapidly aging population. An essential question is that whether the increased life expectancy is due to the growth in healthy or unhealthy years. Healthy life expectancy (HLE) represents the expected number of healthy years of remaining lifetime a member of the life table would experience. After Sullivan (1971) published the method for calculating healthy life expectancy under a period life table, many researchers applied this method and developed its extension, for example, Molla, Wagener, and Madans (2001), Imai and Soneji (2007), Manton, Gu, and Lamb (2006), and many others.

2.1 Deriving Life Expectancy and Healthy Life Expectancy

Theoretically, a real or a hypothetical cohort mortality, which can be considered as a continuous-time process, is determined by the hazard function $\mu(x, y)$, denoting the instantaneous rate of mortality at a given age $x \in [0, \infty]$ for a cohort born at time y. In the age-continuous context, life expectancy of an individual at age x who is born at time y, represented by $e(x, y)$, can be derived given the harzard function $\mu(x, y)$. Let $l(0, y)$ be the total number alive of newborns for this cohort, as the hypothetical cohort that experiences the current observed cross-sectional mortality rates, the number of people survived at age x is

$$l(x, y) = l(0, y) \exp[-\int_0^x \mu(\tau, y) d\tau]. \tag{1}$$

$l(x, y)$ is equivalent with the survival function of this cohort if we normalize $l(0, y)$ to be 1. Then life expectancy, $e(x, y)$ can be computed as

$$e(x, y) = \frac{1}{l(x, y)} \int_x^\infty l(\tau, y) d\tau. \tag{2}$$

Sullivan (1971) employed a relatively simple modification of the conventional life table model to compute the expected duration of certain defined conditions of interest among the living population. For example, the expected remaining healthy lived years for an individual, which is the so called healthy life expectancy (HLE). A variable called *disability prevalence ratio*, denoted by $\pi(x, y)$, is commonly used in the literature about Sullivan's method. $\pi(x, y)$ is the proportion disabled at age x for the cohort born at time y. That is, given that an individual of this cohort who survived up to age x, the conditional probability that he/she is disabled at age x.

In this thesis, $\pi(x, y)$ is defined as the *Health Status Index* (HSI), which reflects the proportion of population in bad health for a cohort that has birth year y at age x. Consequently, the number of survivors who are healthy at age x is $[1 - \pi(x, y)]l(x, y)$. Healthy life expectancy $e^H(e, y)$ in turn can be computed as

$$e^H(x, y) = \frac{1}{l(x, y)} \int_x^\infty [1 - \pi(\tau, y)]l(\tau, y)d\tau. \tag{3}$$

In practice, discrete data is usually adopted to construct approximations of the continuous-time life table functions. I will first illustrate the traditional Sullivan's method without the time component in a period life table within the discrete data framework, and then address a cohort life table by including the time component, which can determine life expectancy for specific cohort.

2.2 Period Life Table

Sullivan's approach of computing healthy life expectancy is derived from a period life table based on discrete data. A general setting of life expectancy analysis based on a period life table will be described in this section, and a specific setting adopted by this paper will be specified in section 2.4. Let n_x denote the length of an age interval starting at age $x \in \mathcal{A}$. $\mathcal{A}$ is the set of the starting ages for the age intervals of a period life table. Except the oldest age interval $[\omega, \infty)$ which starts at age ω, all the other age intervals have the same length ($n_x = n$). Molla, Wagener, and Madans (2001) argued that the age beginning at the oldest age interval does not have any effect on a life table being constructed. When $n = 1$, a period life table is called unabridged, and it is said to be abridged if $n > 1$.

Sullivan's computations of the expectation for healthy life is based on the stationarity assumptions of the population, which are illustrated in detail by Chiang (1984) and Preston, Heuveline, and Guillot (2001) as follows,

1. The age-specific hazard rate is constant over time, i.e. $\mu(x, y) = \mu(x)$.
2. The birth rate is constant over time
3. The net migration rates at all ages are zero.

The stationarity assumptions indicate the following,

1. The survival function is constant over time, i.e. $l(x, y) = l(x)$.
2. The raw death rate equals the raw birth rate.
3. The total size of the hypothetical cohort is assumed to remain constant over time.
4. The age distribution in any interval $[x, x + n_x)$ of the hypothetical cohort is constant over time and is proportional to the survival function. That is, for age $s \in [x, x + n_x)$, the density of the age distribution is $\frac{l(s)}{\int_x^{x+n_x} l(\tau)d\tau}$.

Thus, the age-specific mortality rate, which is denoted by ${}_{n_x}M_x$, can be written as,

$$ {}_{n_x}M_x = \frac{\int_x^{x+n_x} l(\tau)\mu(\tau)d\tau}{\int_x^{x+n_x} l(\tau)d\tau}. \tag{4} $$

Note that the time component is not modeled in Sullivan's method because of stationarity.

In the age-continuous context, notations like $q(x)$, $l(x)$, $e(x)$ etc. are commonly used, whereas for age-discrete calculations, notations like q_x, l_x, e_x, etc. are adopted in common demographic notation.

The starting point of creating a period life table in the discrete context is to include the total number of person-years in a population over a calendar year, which is the so called exposure-to-risk ${}_{n_x}E_x$, and the total number of deaths within an entire year ${}_{n_x}D_x$ for the interval $[x, x+n_x)$, where the prescripts indicate the length of the interval under consideration. The central death rate for this interval, denoted by ${}_{n_x}m_x$, can be written as,

$$ {}_{n_x}m_x = \frac{{}_{n_x}D_x}{{}_{n_x}E_x}. \tag{5} $$

${}_{n_x}m_x$ is an estimator of ${}_{n_x}M_x$ in (4), because, ${}_{n_x}E_x$ and ${}_{n_x}D_x$ are usually obtained from the census data and vital statistics in practice, and they are very large, see Imai and Soneji (2007).

Then, ${}_{n_x}q_x$, representing the conditional probability of death within an age interval with length n_x, given that an individual of the hypothetical cohort survived up to age x, can be calculated as, (see Molla, Wagener, and Madans (2001))

$$ {}_{n_x}q_x = \frac{n_{x\,n_x}m_x}{1 + n_x(1 -_{n_x} a_x)_{n_x}m_x}, \tag{6} $$

where ${}_{n_x}a_x$ is the average proportion of years lived in the age interval $[x, x+n_x)$ among those who are alive at age x but die within the interval, and can be obtained from complete life tables. Hence, l_{x+n_x}, the number of alive at age $x+n_x$, is calculated by multiplying l_x, the number of survivors at age x, by the probability of surviving from age x to $x+n_x$, $(1 -_{n_x} q_x)$. That is,

$$ l_{x+n_x} = l_x(1 -_{n_x} q_x). \tag{7} $$

The total number of person-years lived in this interval is then given by

$$ {}_{n_x}L_x = n_x l_{x+n_x} + l_{x\,n_x}q_{x\,n_x}a_x, \tag{8} $$

where $l_{x\,n_x}q_x$ means the proportion who die in the interval contributes ${}_{n_x}a_x$ years on average. Within this framework, life expectancy at age x can be written as

$$ e_x = \frac{1}{l_x}\sum_{i\in\mathcal{A}_\S} {}_{n_i}L_i, \tag{9} $$

where $\mathcal{A}_x = \{i \in \mathcal{A} : x \le i\}$.

Imai and Soneji (2007) showed that under the stationarity assumptions, e_x calculated from the discrete data equals $e(x)$ in the theoretical definition (2). This is because, l_x used in discrete setting and $l(x)$, see (7), used in continuous setting both refer to the proportion alive at exact age x, thus they are numerically identical. Moreover, in the continuous context,

$$ {}_{n_x}q(x) = \frac{\int_x^{x+n_x} l(\tau)\mu(\tau)d\tau}{l(x)}, \tag{10} $$

$$ {}_{n_x}a(x) = \frac{\int_x^{x+n_x} l(\tau)\mu(\tau)(\tau - x)d\tau}{\int_x^{x+n_x} l(\tau)\mu(t)d\tau}. \tag{11} $$

Substituting (10) and (11) into (8) and integrating by parts yield

$$ {}_{n_x}L_x = \int_x^{x+n_x} l(\tau)d\tau \tag{12} $$

This proves that e_x equals $e(x)$.

2.3 Healthy Life Expectancy from Sullivan's Method

The life table measure is of great use to estimate the remaining lifetime of a group of persons with a certain age. However, whether the remaining life is in good health is another crucial issue regardless of their ages. By including additional age-specific information of health status into a period life table, Sullivan (1971) suggested a measure to separate the remaining lifetime into a healthy and an unhealthy part. The healthy years that are spent during the whole remaining years of living is the so called healthy life expectancy, and can be estimated from cross-sectional data by

$$ \hat{e}_x^H = \frac{1}{l_x} \sum_{i \in \mathcal{A}} (1 - {}_{n_i}\hat{\pi}_i) {}_{n_i}L_i, \tag{13} $$

Sullivan (1971) originally defined ${}_{n_i}\pi_i$ as the *disability prevalence ratio* and suggested in his paper the following estimator,

$$ {}_{n_i}\hat{\pi}_i = \frac{1}{{}_{n_i}N_i} \sum_{j=1}^{{}_{n_i}N_i} \frac{W_{ij}(t_{ij})}{365}, \tag{14} $$

where $W_{ij}(t_{ij})$ is the self-reported number of days of disability per year for the jth respondent in the interval beginning at age i, and $\hat{e}_x^H$ in (13) corresponds to *disability free life expectancy*. However, Imai and Soneji (2007) showed that it is unlikely to estimate disability free life expectancy without bias using $W_{ij}(t_{ij})$, accordingly to the disability prevalence ratio over the one-year period. Rogers, Rogers, and Belanger (1990) also proved Sullivan's method actually underestimates disability free life expectancy because of the bias in the estimation of the disability prevalence.

Hence, Imai and Soneji (2007) proposed ${}_{n_i}\hat{\pi}_i$ is the sample fraction of the disabled among the survey respondents within the age interval $[i, i + n_i)$. Most of the applications, including Imai and Soneji (2007) use the following measure to estimate ${}_{n_i}\pi_i$

$$ {}_{n_i}\hat{\pi}_i = \frac{1}{{}_{n_i}N_i} \sum_{j=1}^{{}_{n_i}N_i} Y_{ij}(t_{ij}), \tag{15} $$

where ${}_{n_i}N_i$ denotes the total number of the survey respondents in the age interval $[i, i + n_i)$, and $Y_{ij}(t_{ij})$ is the disability indicator for the jth respondent of that interval whose age is $t_{ij} \in [i, i + n_i)$ at the time of the survey. Most of the literature adopts (15) as the estimate of ${}_{n_i}\pi_i$. Imai and Soneji (2007) proved that by incorporating only one additional stationarity assumption, which is the age-specific disability prevalence ratio is constant over time, i.e. $\pi(x, y) = \pi(x)$ for all y, Sullivan's estimator is unbiased and consistent, and the standard variance estimator is consistent and approximately unbiased. Imai and Soneji (2007) pointed out that the estimator ${}_{n_i}\hat{\pi}_i$ from (15) also can be computed as a weighted average with appropriate sampling weights.

Differently to the current literature, measures of health status other than disability are used in this thesis to refine the decomposition of life expectancy. $Y_{ij}(t_{ij})$ in (15), is redefined as the

indicator of bad health of the jth respondent of that interval whose age is $t_{ij} \in [i, i+n_i)$ at the time of the survey. The corresponding ${}_{n_i}\pi_i$, which reflects the proportion of the population in bad health is called *health status index*.

2.4 Cohort Life Table

However, including Sullivan (1971), many researchers point out that since the age-specific rates may change considerably over the lifespan of any real birth cohort, expectations based on a period life table solely may not reflect accurately the life experience of infants born in any specific period. Imai and Soneji (2007) proved that life expectancy can be estimated without stationarity and other assumptions by using a cohort life table. The estimation still remains unbiased with consecutive cross-sectional data. For this reason, life expectancy will be created using a cohort life table in this thesis based on the consecutive cross-sectional surveys, which are often easier to obtain, to construct a cohort life table. The age interval is chosen to be one year, that is $n_x = n = 1$. Therefore, for notational simplification, the prescripts n_x for the corresponding notations are omitted. In summary, the procedures of constructing a cohort life table and calculating life expectancy from the consecutive cross-sectional data are as follows. Note that explicit reference to the year of birth y is trivially given by $t = y + x$.

1. First observe the total number of death $D_{x,t}$, and the exposure-to-risk $E_{x,t}$ to calculate the central death rate
$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}. \tag{16}$$
2. Assume $a_{x,t} = 0$ and choose $n_x = 1$, according to (6), the conditional probability of death for this cohort is
$$q_{x,t} = \frac{m_{x,t}}{1 + m_{x,t}} \tag{17}$$
and the survival probability $p_{x,t} = 1 - q_{x,t}$ follows.
3. The quantities $l_{x,t}$ and $L_{x,t}$ are equal in value in this framework according to (8), if $a_{x,t} = 0$, $n_x = 1$, and we normalize $l_{0,t} = 1$,
$$l_{x,t} = L_{x,t} = l_{x-1,t-1} \times p_{x-1,t-1} = p_{0,t-x} \times \ldots \times p_{x-1,t-1}.$$
4. Consequently, life expectancy in a cohort life table can be estimated as follows,
$$\hat{e}_{x,t} \quad = \quad \frac{1}{l_{x,t}} \sum_{i \in \mathcal{A}_x} L_{i,t}. \tag{18}$$

2.5 Healthy Life Expectancy Using Cohort Life Table

Sullivan's healthy life expectancy can be estimated in an unbiased and consistent way without stationarity assumptions by using the consecutive cross-sectional health data based on a cohort life table. Healthy life expectancy is derived by involving the health status index for the cohort age age x of year t, $\hat{\pi}_{x,t}$ into (18),

$$\hat{e}^H_{x,t} = \frac{1}{l_{x,t}} \sum_{i \in \mathcal{A}_\S} (1 - \hat{\pi}_{i,t}) L_{i,t}. \tag{19}$$

where $\hat{\pi}_{x,t}$ can be calculated from the health surveys defined analogously as (15),

$$\hat{\pi}_{x,t} = \frac{1}{N_{x,t}} \sum_{j=1}^{N_{x,t}} Y_{ij}(t_{ij}). \tag{20}$$

where $Y_{ij}(t_{ij})$ is the indicator of bad health of the jth respondent of that interval whose age is $t_{ij} \in [i, i + n_i)$ at the time of the survey.

3 Modeling Future Mortality and Health Status

The current literature incorporates the health status into a life table to derive healthy life expectancy. However, a small part of the literature directly examines health changes in individuals over time or include the time component. Therefore, the analysis constructed in this thesis will include the time component t, which corresponds to a cohort life table. And, to construct a cohort life table from the consecutive cross-sectional data, it is necessary to model and project the mortality process, to obtain the corresponding life expectancy. Moreover, a proper model is needed to model the dynamic changes of the health status process and project the future trend to derive healthy life expectancy. The Lee-Carter model is thus adopted for modeling the health process. In this section, first the Lee-Carter model will be illustrated in the mortality context, then how to apply the Lee-Carter model on describing the stochastic changes of the health status index process will be addressed.

3.1 The Lee-Carter Model

Lee and Carter (1992) proposed a simple model for describing the changes in total mortality as a function of a single time parameter, κ_t. This parsimonious dynamic mortality model turned out to perform quite well for the U.S. data. Let

$$m_{x,t}, x = x_1, x_2, \ldots, x_k, t = t_1, t_2, \ldots, t_n,$$

denote the central death rate for age x at time t. The Lee-Carter model postulates the following log-bilinear relationship:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{21}$$

where κ_t is a time-dependent univariate mortality index, which represents the change in the level of mortality over time. α_x describes the age-pattern of mortality averaged over time, while β_x describes the age-specific deviations from the averaged pattern when κ_t varies. The $\epsilon_{x,t}$ (white noise) denotes the error term, with mean 0 and variance $\sigma^2_{\epsilon,x}$, reflecting particular age-specific historical influences not captured by the model.

β_x and κ_t cannot be uniquely identified, because one of these two elements could be multiplied by a constant while the other one is divided by the same constant without altering the predicted values given by the model. Hence, Lee and Carter (1992) proposed the normalization constraints,

$$\sum_t \kappa_t = 0, \sum_x \beta_x = 1. \tag{22}$$

The first one implies that for each x the estimate for α_x will be an average of the log-central rate of morality over calendar years. The second constraint is to uniquely identify β_x and κ_t. Cairns (2007) argued that the first constraint is natural, but not for the second one. However, different choices of the second constraint has no impact on the quality of the fit, or the mortality forecasts. Researchers also propose other constraints, for instance, Wilmoth (1993) adopted $\sum_t \kappa_t = 0$ and $\sum_x \beta_x^2 = 1$.

Since there is no observable variable on the right-hand side of (21), the model cannot be fitted by conventional regression methods. Lee and Carter (1992) proposed a singular value decomposition (SVD) method to find a least squares solution. Let the central death rate $m_{x,t}$, also denote the observed (raw) mortality rate, the model fitting procedures have three steps as follows,

1. Specifically, parameters α_x, β_x and κ_t are estimated by minimizing

$$\mathcal{F}_{LS}(\alpha,\beta,\kappa)=\sum_{x=x_1}^{x_k}\sum_{t=t_1}^{t_n}(\ln(m)_{x,t}-\alpha_x-\beta_x\kappa_t)^2. \tag{23}$$

   By taking the partial derivative of $\mathcal{F}_{LS}(\alpha,\beta,\kappa)$ with respect to α and setting it to be 0, we have

$$\hat{\alpha}_x=\frac{1}{t_n-t_1+1}\sum_{t=t_1}^{t_n}\ln(m_{x,t}), \tag{24}$$

   since $\sum_{t=t_1}^{t_n}\kappa_t=0$.
2. Fit the model (21) to a matrix of observed mortality rate $m_{x,t}$ using singular value decomposition. The estimated κ_t and β_x are the respective first right and first left singular vectors in the SVD of the matrix $\{\ln(m_{x,t})-\hat{\alpha}_x\}$.
3. Finally, κ_t are reestimated by fixing $\hat{\alpha}_x$ and $\hat{\beta}_x$ so that the actual total observed deaths equal the total expected deaths for each year t.

The adjustment of each κ_t gives greater weight to ages at which numbers of deaths are large. By allowing for the constraints (22), the number of free parameters is $2k+n-2$. In order to avoid taking logarithms of zeros, Wilmoth (1993) proposed a weighted SVD by replacing the objective function (23) with

$$\mathcal{F}_{WLS}(\alpha,\beta,\kappa)=\sum_{x=x_1}^{x_k}\sum_{t=t_1}^{t_n}\omega_{x,t}(\ln(m)_{x,t}-\alpha_x-\beta_x\kappa_t)^2, \tag{25}$$

where the weight $\omega_{x,t}$ equals to the observed number of deaths in each cell of the data matrix empirically. A weighted SVD is typically designed for estimating a variable without big sample size, which easily encounters the zero elements in the selected sample.

Alternatively, to avoid the Singular Value Decomposition (SVD), parameters α_x, β_x and κ_t can also be estimated from the Newton-Raphson recursive procedures; a detailed description is illustrated by Pitacco, Denuit, Haberman, and Olivieri (2009). In the Newton-Raphson procedures, first obtain the partial derivatives of $\mathcal{F}_{LS}(\alpha,\beta,\kappa)$ given in (23) with respect to α_x, β_x, and κ_t, and set the partial derivatives equal to 0 respectively,

$$0 = \sum_{t=t_1}^{t_n}(\ln(m)_{x,t}-\alpha_x-\beta_x\kappa_t), \tag{26}$$

$$0 = \sum_{x=x_1}^{x_k}\beta_x(\ln(m)_{x,t}-\alpha_x-\beta_x\kappa_t), \tag{27}$$

$$0 = \sum_{t=t_1}^{t_n}\kappa_t(\ln(m)_{x,t}-\alpha_x-\beta_x\kappa_t). \tag{28}$$

the estimate for α_x is given by (24), then estimated $\hat{\beta}_x$ and $\hat{\kappa}_t$ are updated iteratively by the univariate Newton-Raphson scheme. For example, in the rth iteration, the recursive relations are specified as follows,

$$\hat{\kappa}_x^{(r+1)} = \hat{\kappa}_t^{(r)}+\frac{\sum_{x=x_1}^{x_k}\hat{\beta}_x^{(r)}(\ln(m)_{x,t}-\hat{\alpha}_x-\hat{\beta}_x^{(r)}\hat{\kappa}_t^{(r)})}{\sum_{x=x_1}^{x_k}(\hat{\beta}_x^{(r)})^2}, \tag{29}$$

$$\hat{\beta}_x^{(r+1)} = \hat{\beta}_x^{(r)}+\frac{\sum_{t=t_1}^{t_r}\hat{\kappa}_t^{(r+1)}(\ln(m)_{x,t}-\hat{\alpha}_x-\hat{\beta}_x^{(r)}\hat{\kappa}_t^{(r+1)})}{\sum_{t=t_1}^{t_n}(\hat{\kappa}_t^{(r+1)})^2.} \tag{30}$$

Finally, these parameters are adjusted by the identifiability constraints (22), and $\hat{\kappa}_t^{(r+1)}$ are further adjusted by fitting the total observed deaths to the total expected deaths for each year t. This iteration will be proceeded R times until we get the smallest difference between the estimated deaths and the observed deaths.

There is a wide class of generalized, parametric, non-linear extended models based on the simple Lee-Carter framework. For instance, Renshaw and Haberman (2003b) included the first two sets of SVD vectors in the estimation and forecast, rather than just the first such set of vectors. Renshaw and Haberman (2003c) argued that using the mortality reduction factors are very important for capturing and projecting historic mortality trends. Renshaw and Haberman (2003a) introduced a generalized linear modeling technology as a parallel methodology with the Lee-Carter model and compared the two models in terms of structure and assumption. Later on, to capture the age-period cohort effect, Renshaw and Haberman (2005) incorporated the age-period cohort effect as an additional variable into the Lee-Carter model to improve the mortality projection. Cairns, Blake, and Dowd (2006) introduced a two-factor stochastic model for the development of mortality through time. The first factor affects mortality-rate dynamics at all ages in the same way, whereas the second factor affects mortality-rate dynamics at higher ages much more than at lower ages.

3.2 Forecasting by the Lee-Carter Approach

The Lee-Carter model uses the Box-Jenkins method to identify and estimate the dynamics of the latent factor κ_t within an ARIMA time series model. Although, this is not necessarily a linear relationship, Lee and Carter (1992) and most of the other literature, including Tuljapurkar, Li, and Boe (2000) concluded that the dynamics of κ_t can be described as a random walk with drift μ. This ARIMA(0,1,0) time series model is,

$$\kappa_t = \mu + \kappa_{t-1} + e_t, \tag{31}$$

where the innovation e_t is assumed to follow a normal distribution with mean 0 and variance σ_e^2. Then, the m ahead point forecast through an ARIMA(0,1,0) model can be derived as follows,

$$\tilde{\kappa}_m = \kappa_1 + (m-1)\mu. \tag{32}$$

The forecasts of κ_t in turn yield projected age-specific mortality rates,

$$\ln(\tilde{m}_{x,m}) = \hat{\alpha}_x + \hat{\beta}_x \tilde{\kappa}_m. \tag{33}$$

Life expectancy can be computed from a cohort life table based on the projected $\ln(\tilde{m}_{x,m})$ corresponding to the projected $\tilde{\kappa}_m$. The maximum likelihood estimator of μ and σ_e^2 in (31) are the sample mean and variance of the first order integration of κ_t; these are

$$\hat{\mu} = \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} (\hat{\kappa}_t - \hat{\kappa}_{t-1}) = \frac{\hat{\kappa}_{t_n} - \hat{\kappa}_{t_1}}{t_n - t_1}, \tag{34}$$

$$\hat{\sigma}_e^2 = \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} (\hat{\kappa}_t - \hat{\kappa}_{t-1} - \hat{\mu})^2. \tag{35}$$

Using $\hat{\sigma}_e^2$, we can construct the confidence interval for $\tilde{\kappa}_t$.

Note that $m_{x,t}$ is modeled as a stochastic process, which is driven by the stochastic process κ_t, from which interval estimates can be computed for the projected values of mortality rates. The corresponding variance of the projected logarithm of the mortality rate is

$$Var(\ln(\tilde{m}_{x,t+m})) = \hat{\beta}_x^2 m \hat{\sigma}_e^2. \tag{36}$$

It is worth mentioning that the above point forecasts of κ_t and $m_{x,t}$ from (32) and (33) are derived without taking into account any stochastic development of the process, which is called *process risk*, or the uncertainty caused by the inaccuracy of the estimated parameters, which is called *parameter risk*. Later on in the empirical analysis, these two risks will be included using the simulation method to stochastically forecast the mortality.

3.3 Health Modeling

It should be noted that the underlying assumption of $\pi_{x,y}$ used when calculating healthy life expectancy is that $\pi_{x,y}$ does not change by a large margin across different cohorts. Researchers also start to estimate $\pi_{x,y}$ using stochastic models, for example, the generalized additive models (GAMs) proposed by Hastie and Tibshirani (1986). This thesis proposes to capture the random element in the stochastic development of health process using the Lee-Carter approach, besides its wide application on mortality rates. This is helpful to model the health status process and capture its stochastic change in the future, which is beneficial to create healthy life expectancy based on a cohort life table. Hence, $\ln(m_{x,t})$ in the Lee-Cater approach (21), is replaced by the health status index $\pi_{x,t}$ as follows,

$$\pi_{x,t} = \alpha_x^H + \beta_x^H \kappa_t^H + \epsilon_{x,t}^H. \tag{37}$$

Alternatively, one can use the logit transformation of $\pi_{x,t}$ in the Lee-Carter model as well,

$$logit(\pi_{x,t}) = \ln(\frac{\pi_{x,t}}{1-\pi_{x,t}}) = \alpha_x^{H'} + \beta_x^{H'} \kappa_t^{H'} + \epsilon_{x,t}^{H'}. \tag{38}$$

Projecting $\pi_{x,t}$ and $logit(\pi_{x,t})$ follows the same measure as (32) and (33), in which estimated parameters in the mortality context are replaced by the parameter estimates in the health context from (37) and (38). Moreover, confidence intervals of projected $\pi_{x,t}$ and $logit(\pi_{x,t})$ can be constructed under the same method as described in section 3.2 for mortality rates. Later on, process risk and parameter risk will be included in the empirical analysis for the health process as well.

4 Data and Empirical Analysis

4.1 Data

The empirical analysis in this thesis is based on the consecutive annual cross-sectional mortality rate and health status data from 1972 to 2006 in the United Sates. The mortality data is obtained from the Human Mortality Database[1] (HMD), which contains detailed population and mortality of the U.S.. The health status data is obtained from the Integrated Health Interview Series (IHIS), which provides the consecutive cross-sectional data which is the harmonized data and documentation for the U.S. National Health Interview Survey (NHIS).

4.1.1 Mortality Data

The mortality data obtained from the Human Mortality Database constitutes the number of deaths, $D_{x,t}$ and the exposure-to-risk, $E_{x,t}$ at age x of year t, from which the raw (observed) mortality rate is computed according to (16). Figure 1 shows the raw mortality rate from 1972 to 2006 by gender relative to average mortality rates between 1972-1976 of ages 25, 45, 65, and 85. In line with the previous literature (see, for example, Cairns (2007)), the

[1]The website of Human Mortality Database is http://www.mortality.org/

relative raw mortality rates exhibit a downward trend over time at different ages, and have been erratic. Figure 2 is the logarithmic mortality rate across different ages and years. It provides a general impression how mortality rates of different ages are related. Given a year, for example, 2000, we can determine the logit mortality rate, that is $\ln(\frac{m_{x,t}}{1-m_{x,t}})$ with $t = 2000$ as plotted in Figure 3. We can see that in the elderly age interval, for example, 40-100, there is a reasonable linear relationship with age x for both genders.

4.1.2 Health Data

Besides life expectancy at a certain age, the remaining years of life that a group of people can expect to live in good health is an important part. Consequently, first how to define good health becomes an important argument in the literature. Many researchers refer to the Grade of Membership (GoM) to divide the health status into different categories, using a variety of variables. However, the self-reported health data, although being a set of subjective data, is also very valuable to be taken into account. Since a person who is willing to work longer must perceive himself (herself) healthy enough to do so, which is important for issues like increasing the retirement age for the social security and pension funds, etc.. To this extent, how people themselves perceive their health status is a very important way to determine the health status. The IHIS provides the integrated self-reported health status of surveyed individuals from 1972 to 2006 and it rates an individual's general health. The self-reported health is obtained by the person in question or evaluated by a family member on a four-point scale (excellent, good, fair, or poor) for 1972-81 or a five-point scale (excellent, very good, good, fair, or poor) from 1982 until now, ranging from "excellent" to "poor" in general. One way to define the health status index is that people are deemed to be healthy unless they report "poor", and the health status index $\pi_{x,t}$ can be estimated by equation (20), where $Y_{xj,t}(t_{xj,t})$ is the indicator that the respondent reports "poor" health; alternatively, persons who rate their health status better than "fair" are deemed to be healthy, and $Y_{xj,t}(t_{xj,t})$ in (20) becomes the indicator that the respondent reports either "poor" or "fair" health.

Figures 4 and 5 are the observed health status index (HSI) by gender that cut at "poor", whereas Figures 6 and 7 plot the HSI that cut at "fair". They show that, generally, elderly people are less healthy than younger people. Figures 5 and 7 provide general impressions how the health status index in level format of different ages are related.

One commonly used transformation of proportion data is the logit transformation (see, equation (38)). The sample descriptions for logit health status index that cut at "poor" and "fair" health, respectively, are given by Figures 8 to 11.

Note that since the sampling scales are changed from 1982 forward, it deserves a careful attention in the estimation analysis. We also can see that a jump happens around 1982 from the sample descriptive figures. The IHIS reports that the relative frequency of responses more favorable than "fair", combining "excellent," "very good," and "good" versus combining "excellent" and "good" is similar before and after 1982. Another irregular movement happens around 1997. This may be because prior to 1997, all persons for whom health status information was unavailable are grouped together and coded as "Unknown". Starting in 1997, the reason why this information was unavailable is specified in detail, which may affect the responses of the survey respondents.

4.2 Empirical Analysis Using the Whole Sample

4.2.1 Mortality Estimation and Life Expectancy

The Lee-Carter Model is first applied to annual mortality rates by gender during the period $\mathcal{T} = 1972, \ldots, 2006$. The sample ages $x \in \mathcal{A}$, where $\mathcal{A} = 0, 1, \ldots, 110$ is the set of starting

ages for the age intervals, and 110 is the starting age for the oldest open age group $[110, \infty)$. The Newton-Raphson recursive procedures introduced in section 3.1 are employed to estimate the Lee-Carter model. The estimated $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$ are plotted in Figures 12 to 14 in the appendix. Figure 12 shows that, generally speaking, males have higher average mortality rates than females in the selected sample. However, males' mortality rates may decrease faster than females', which is indicated by the clear downward trend of $\hat{\kappa}_t$ in Figure 14. Figure 15 shows the in-sample estimated logarithm of the mortality rate, which is a rectangular matrix $\{\ln(m_{x,t})\}, x \in \mathcal{A}, t \in \mathcal{T}$ with dimension 111×35. The estimates are consistent with the observed behavior of mortality rates in Figure 2.

In practice, there are likely irregularities caused by sampling errors, researchers turn to statistical techniques to smooth the estimated parameters and produce a regular progression, see Currie, Durban, and Eilers (2004) and Kirkby and Currie (2010). In this analysis, in order to avoid the erratic behavior, it is necessary to smooth $\hat{\beta}_x$. A widely used method, the B-spline method is used for smoothing $\hat{\beta}_x$, see Renshaw and Haberman (2003b) for instance. For specific details refer to Pitacco, Denuit, Haberman, and Olivieri (2009) (pp. 69-72). The smoothed $\hat{\beta}_x$ is shown in Figure 13, and the smoothed logarithmic mortality rates using smoothed $\hat{\beta}_x$ are presented in Figure 16.

According to the procedures of creating a cohort life table and calculating the expected remaining lifetime of an individual discovered in section 2.4, to forecast life expectancy 20 years ahead from 2006, we need to project κ_t and the corresponding mortality rate $(m_{x,t})$ 130 years ahead, which can be done by following (32) and (33) in section 3.2. This is called the best forecast or point forecast since it does not take any risk into account. Figure 17 shows the forecasted logarithm of mortality rates across ages and over time. Figures 18 and 19 present for a certain age at 65, the 130 years ahead forecasts of κ_t and the logarithm of mortality rates by gender. The 95% forecasting intervals in both cases derived by using the volatility of $\hat{\kappa}_t$ from (35) and the volatility of projected $\ln(m_{x,t})$ from (36) in section 3.2, are also plotted in the figures. These figures show a clear downward trend of mortality rates both in-sample and out-of-sample, among which males' downward trend is steeper than females'.

It's worth mentioning that the projected mortality rate is a rectangular matrix $\{\tilde{m}_{x,t}\}$, with dimension 111×165. Based on the projected $\{\tilde{m}_{x,t}\}$ we can then follow procedures 2, 3 and 4 in section (2.4) to calculate the conditional probability of death, $q_{x,t}$, from where $l_{x,t}$, the hypnotical cohort, and $L_{x,t}$, concerning the total number of person years at age x in year t, can be derived as the products of the diagonal of matrix $\{q_{x,t}\}$, i.e.

$$1 - q_{0,t-x}, 1 - q_{1,t-x+1}, ..., 1 - q_{x-1,t-1}.$$

Consequently, a cohort life table is obtained from the diagonals of a projection matrix $\{q_{x,t}\}$. From (18), $l_{x,t}$ and $L_{x,t}$ accordingly provide the point estimates and forecasts of life expectancy, which are shown in Figure 36.

4.2.2 Risks

Two main sources of risks in the mortality projection are considered in this section. Firstly, since the mortality rate is modeled as a random process, there is *process risk*. Second due to an inaccurate assessment of the relevant parameters, there exists *parameter risk*.

We can include process risk and parameter risk into the mortality projection by means of simulation techniques. I choose 2,000 times simulations in the mortality projection. In this analysis, process risk is generated by e_t in (31) and $\epsilon_{x,t}$ in (21). Under the assumption that e_t is normally distributed with mean 0 and variance σ_e^2, e_t can be generated from a normal distribution with mean 0 and variance $\hat{\sigma}_e^2$ in each simulation, where $\hat{\sigma}_e^2$ is the estimate of σ_e^2 by (35). Moreover, residuals $\epsilon_{x,t}$ in (21) are also assumed to be normally distributed with

mean 0 and age-specific variance $\sigma^2_{\epsilon,x}$, where $\sigma^2_{\epsilon,x}$ can be estimated as follows,

$$\hat{\sigma}^2_{\epsilon,x} = \frac{1}{T}\sum_{t_1}^{t_n}(y_{x,t} - \hat{y}_{x,t})^2. \tag{39}$$

In this equation, $y_{x,t}$ is the logarithm of raw central mortality rate at age x of year t, $\ln(m_{x,t})$; and $\hat{y}_{x,t} = \hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t$ is the smoothed estimated logarithm of the mortality rate. Hence, each simulation generates a $\epsilon_{x,t}$ from a normal distribution with mean 0 and variance $\hat{\sigma}^2_{\epsilon,x}$.

Next, parameter risk considering μ and σ^2_e in ARIMA (0,1,0) process of κ_t is included, see (31), which can be rewritten as

$$\Delta\kappa_t = \mu + e_t \tag{40}$$

where $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$. μ and σ^2_e are estimated from (34) and (35). Typically, we can either apply the standard central limit theorem to derive the asymptotic distribution of $\hat{\mu}$ and $\hat{\sigma}_e$, or employ the Bayesian method by assuming prior distributions for the parameter sets (see, Cairns, Blake, and Dowd (2006)). This analysis uses the first approach by employing the standard central limit theorem, we have

$$\sqrt{T}(\hat{\mu} - \mu) \underset{T\to\infty}{\longrightarrow} \mathcal{N}(0, \theta),$$

where $T = t_n - t_1$ in (34) or (35). θ can be estimated as follows

$$\hat{\theta} = TVar(\hat{\mu}) = TVar(\frac{1}{T}\sum_t \Delta\kappa_t) = Var(\Delta\kappa_t) = \hat{\sigma}^2_e.$$

Since $s^2 = \frac{1}{T-1}\sum_t(\Delta\kappa_t - \hat{\mu})^2$ is an unbiased estimator of σ^2_e and $\frac{s^2}{\sigma^2_e}(T-1) \sim \chi^2(T-1)$, it follows that $Var(s^2) = 2\frac{\sigma^4_e}{T-1}$, which leads to

$$\sqrt{T}(\hat{\sigma}^2_e - \sigma^2_e) \underset{T\to\infty}{\simeq} \sqrt{T}(s^2 - \sigma^2_e) \underset{T\to\infty}{\longrightarrow} \mathcal{N}(0, 2\sigma^4_e).$$

Therefore, the asymptotic distribution of $\hat{\mu}$ and $\hat{\sigma}_e$ under the central limit theorem is

$$\sqrt{T}(\begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2_e \end{bmatrix} - \begin{bmatrix} \mu \\ \sigma^2_e \end{bmatrix}) \underset{T\to\infty}{\longrightarrow} \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_e & 0 \\ 0 & 2\sigma^4_e \end{bmatrix})$$

The off diagonal components turn out to be equal to 0, see, for example, Hamilton (1994) (pp. 298-302). Parameters $\hat{\mu}$ and $\hat{\sigma}^2_e$ used in the mortality projection are in turn drawn from their distributions in each simulation when including parameter risk.

In this study, I either include only process risk caused by e_t, or both process and parameter risks caused by e_t, $\epsilon_{x,t}$, $\hat{\mu}$ and $\hat{\sigma}^2_e$. It is worth mentioning that, when including multiple risks, every risk factor is generated together in one simulation. Expected simulated life expectancy is the average of 2,000 simulated paths. By sorting the 2,000 simulated $\hat{LE}_{x,t}$, we can derive the 95% confidence interval of expected simulated $\hat{LE}_{x,t}$ with lower and higher bounds to be the 2.5% and 92.5% quantiles. Figure 39 presents the comparison of point estimates of LE and expected simulated LE with its 95% confidence interval at a certain age $x = 65$. It shows that expected simulated LE with different risks and the point estimates are very close to each other in value for both females and males; there is almost no deviation from the point estimates and forecasts when including parameter risk and process risk. As expected, the uncertainty of simulated LE with both process and parameter risks are larger than with process risk only, which is reflected by a wider confidence interval when including two risks. Males possess

lower LE than females and wider confidence interval in general, but their life expectancies increase faster than females' over time.

In order to avoid the simulation risk when simulating life expectancy, it is necessary to determine the simulation errors. Let $\{LE_1, \ldots, LE_S\}$ denote life expectancy simulated in S simulations, with mean μ_{LE}. μ_{LE} can be estimated by the simulation average, that is $\hat{\mu}_{LE} = \frac{1}{S}\sum_s LE_s$. According to the central limit theorem, when S goes to infinity,

$$\sqrt{S}(\hat{\mu}_{LE} - \mu_{LE}) \underset{S\to\infty}{\longrightarrow} \mathcal{N}(0, \sigma^2_{LE}),$$

where σ^2_{LE} is the variance of $\{LE_1, \ldots, LE_S\}$, and can be estimated as,

$$\hat{\sigma}^2_{LE} = \frac{1}{S}\sum_s (LE_s - \hat{\mu}_{LE})^2.$$

As a consequence,

$$\hat{\mu}_{LE} \approx \mathcal{N}(\mu_{LE}, \frac{1}{S}\sum_s (LE_s - \hat{\mu}_{LE})^2).$$

And $\hat{\sigma}^2_{LE} = \frac{1}{S}\sum_s (LE_s - \hat{\mu}_{LE})^2$ is the corresponding simulation error. The simulation errors of simulated life expectancy are listed in the second column of table 2 in percentage. Their small magnitudes suggest that simulation risk is negligible.

4.2.3 Health Estimation and Healthy Life Expectancy

A cohort life tables is a powerful technique for estimating the remaining years of life of an individual. However, it is essential for examining whether the remaining years of life are healthy or not. Hence, this thesis focuses not only on life expectancy, but also on its healthy years. In this section, Sullivan's method is applied for estimating healthy life expectancy by including the health status index $\pi_{x,y}$. To be precise, the definition of $\pi_{x,y}$ used in this thesis is different as originally defined in Sullivan (1971) or in Molla, Wagener, and Madans (2001), and Imai and Soneji (2007). $Y_{xj,t}(t_{xj,t})$ in (15) is replaced to be the indicator of bad health, then $\pi_{x,y}$ is accordingly called the *health status index* (HSI), see section 2.3. This is a meaningful attempt because people's health status is closely related to their ability to work as argued.

In this section, four different models will be examined and compared, because we can measure $\pi_{x,t}$ from the self-reported health status data in two ways, namely if bad health is defined as "poor" or "poor" plus "fair" health states. Under each HSI definition, $\pi_{x,t}$ can be incorporated into the Lee-Carter model either with level format or logit format, see (37) and (38). Therefore, this offers four different combinations for estimating the health process, that are $\pi_{x,t}$ cutting at "poor" with level format, $\pi_{x,t}$ cutting at "poor" with logit format, the level format of $\pi_{x,t}$ cutting at "fair", and the logit format of $\pi_{x,t}$ cutting at "fair".

The sample ages $x \in \mathcal{A}$, where $\mathcal{A} = 0, 1, \ldots, 85$, and the sample period is from 1972 to 2006. The starting age of the oldest age interval is chosen to be 85 in order to match the sample size of self-reported health over time. As in section 4.2.1 for the mortality estimation, I use the Newton-Raphson recursive procedures illustrated in section 3.1 to estimate the Lee-Carter model for level and logit health status index under different measures. Additionally, both the estimated α^H_x ($\alpha^{H'}_x$) and β^H_x ($\beta^{H'}_x$) are smoothed in order to avoid an erratic progression. Because the sample size of the health status provided by the IHIS is much smaller compared to mortality rates from the Human Mortality Database. Moreover, due to IHIS's yearly budget constraint, the sample size of the health status reduces fiercely in two years, that are 1986 and 1996. These may induce sample errors and lead to the irregular behavior of the HSI.

The estimation results of the health status index under the two definitions with the two formats are shown from Figures 20 to 29 for both females and males. First, Figures 20 and 21 show that people's average health becomes worse as the age increases, no matter using which format or how to measure the HSI. The estimated $\hat{\kappa}_t^H$ ($\hat{\kappa}_t^{H'}$) in Figures 24 and 25 exhibit fluctuating downward trends, where the fluctuations mainly happen around the years 1982, 1986, and 1996. This may be because of an additional category of health status that was included in 1982, and the sample size decreased largely due to the budget constraint of the IHIS in 1986 and 1996. However, due to the very small magnitude of estimated $\hat{\beta}_x^H$ ($\hat{\beta}_x^{H'}$), although we do see trends from $\hat{\kappa}_t^H$ ($\hat{\kappa}_t^{H'}$), but not clearly from the observed HSI (see figures 4, 6, 8, and 10). Figures 26 to 29 show the estimated and smoothed HSI across ages and over years by gender.

To estimate healthy life expectancy and its 20 years' forecasts, we need to project 130 years ahead for the HSI in order to be consistent with a cohort life table constructed in section 4.2.1. Figures 30 and 31 show the forecasted $\pi_{x,t}$ and $logit(\pi_{x,t})$, corresponding to the forecasted κ_t^H and $\kappa_t^{H'}$ in Figures 32 and 33. Figures 34 and 35 plot the forecasted $\pi_{x,t}$ and $logit(\pi_{x,t})$ at age 65 with 95% confidence intervals. Generally, when using $logit(\pi_{x,t})$ that cut at "poor" health, there appears a shaper downward trend compared with cutting at "fair" health, which implies that the proportion of people who have poor health at a certain age and year typically will decrease faster than the proportion with fair health. However, when using the level format, the results are actually reverse.

Next, healthy life expectancy can be computed by including the projected $\pi_{x,t}$ into a cohort life table constructed in section 4.2.1 by using (19). An additional assumption has to be made for the health status index to derive healthy life expectancy. That is persons who are older than 85 are assumed to have the same probability of being in the bad health condition as persons at age 85 of the same year. It should be noted that we are still working on the diagonal elements of the rectangular matrix of $q_{x,t}$ and incorporating $\pi_{x,t}$ correspondingly. The point estimates and 20 years ahead forecasts of healthy life expectancy are presented in Figures 37 and 38 by gender, which show that healthy life expectancies are decreasing over age, but increasing over time.

An intuitive comparison can be found in table 1, which gives examples at ages 0, 25, 45, 65, and 86, and in the years 1985, 1995, 2005, 2015, and 2025. Every last column for "Female" and "Male" lists the Increasing Rate (IR) of LE and HLE over time relative to the ones at 1985. We can see that healthy life expectancy is lower than life expectancy, but increases a bit faster over time. LE and HLE are decreasing over age but increasing over time.

4.2.4 Risks

Process and parameter risks are included into the analysis of healthy life expectancy by means of simulation. 3,000 times simulations are chosen here in order to exclude simulation risk, which can be realized by choosing sufficient number of simulations to keep the small magnitudes of the simulation errors. Figures 42 and 43 present the comparison of expected simulated healthy life expectancies by considering different risks for the four models. Similar to the comparison of life expectancy, including different risks does not really make the expected simulated healthy life expectancy deviate from its point estimates. When using the logit format, the point estimates are slightly higher than the expected simulations. Moreover, comparing healthy life expectancy corresponding to the HSI which takes both fair and poor as bad health, and healthy life expectancy which only treats poor as bad health, the former HLE is lower as expected. This is due to the proportion of people who have better health above "fair" is smaller than the proportion of people who have better health above "poor". Furthermore, the uncertainties of including both process and parameter risks are larger than just

including process risk caused by e_t, which can be seen from the wider confidence intervals provided when including the two risks.

A comparison between expected simulated life expectancies and healthy life expectancies using the different models with only process risk of e_t is presented in Figure 40, in which the uncertainty intervals at 95% are also plotted. It shows that healthy life expectancy is generally lower than life expectancy, but increasing faster with wider confidence intervals. And males' life expectancy and healthy life expectancy increase faster than females' with wider confidence intervals, in line with the results listed in table 3. The same results hold for a comparison between expected simulated life expectancies and healthy life expectancies with both process and parameter risks in Figure 41.

Now, it is natural to come up with a question, which format is better, level or logit? Figures 44 and 45 compare HLEs derived from these formats. It can be seen that, no matter including which risk, HLE derived from logit data are slightly lower than derived from level data, and also increasing a bit slower. However, the uncertainties generated when using the level data are always larger than using the logit data. Table 3 also provides a comparison of expected simulated LE and HLE at the sample age 65, which more intuitively provides the same results as the figures indicate. Overall, healthy life expectancies simulated from the models using the logit data and the level data are very similar as expected, but with different uncertainties. The above results hold for both females and males.

For the sake of comparison, we can also employ the standard mean squared errors, computed for different models, as a criterion. The mean squared error is defined as follows,

$$MSE(\hat{y}) = E[(\hat{y} - y)^2], \tag{41}$$

where y is the observed variable. It can be the observed logarithm of the mortality rate $\ln(m_{x,t})$, or the observed health status index $\pi_{x,t}$, whereas $\hat{y}$ is the corresponding estimated value through the Lee-Carter model after smoothing. The expectation on the right hand side of (41) can be replaced by the sample average of the $(\hat{y} - y)^2$ across ages and over years. It is worth mentioning that when calculating the MSE for the health status index of level or logit format, it is necessary to transform the logit format to the level format, or the other way around, to keep the comparison consistent. The second rows of the last two panels in table 7 list the mean squared errors for the different formats of the HSI in the whole sample analysis. It shows that when only poor health is considered as bad health, different formats of the HSI almost do not affect the model fit. However, when defining both poor and fair health states as bad health, the model using the level format outperforms using the logit transformation.

Moreover, the simulation errors for healthy life expectancy are listed in the last four columns in table 2. Their small magnitudes indicate that simulation risk can be ignored in the analysis of healthy life expectancy as well.

4.3 SubSample Analysis

Although Lee and Carter (1992) addressed that as long as the sample period is more than about 10-20 years, the length of the mortality time series is not crucial. However, Lee and Miller (2001) later obtained better fits by using a calibration period that starts at 1950 instead of 1900 in Lee and Carter (1992). In this analysis, due to an erratic behavior of κ_t^H ($\kappa_t^{H'}$) in the Lee-Carter estimation of the health process when using the whole sample, it should be worthy to try a subsample calibration to obtain a more regular estimate. In the mortality context, Booth, Maindonald, and Smith (2002) designed procedures for selecting an optimal calibration period which identifies the longest period for which the estimated mortality index parameter κ_t is linear. Denuit and Goderniaux (2005) suggested a statistical method that maximizes the adjusted R^2, the classical goodness-of-fit criterion in linear regression, to select

the starting year if κ_t is best approximated by a straight line. Although this thesis does not apply the statistical method, but it obeys the rule proposed by Booth, Maindonald, and Smith (2002) and Denuit and Goderniaux (2005) to choose the optimal calibration period in the health context, by selecting the longest period for the estimated κ_t^H ($\kappa_t^{H'}$) to be linear, which is subjectively observed from the whole sample estimated κ_t^H ($\kappa_t^{H'}$). Accordingly, I choose the sample mortality and health range from 1982 to 2006 at older ages, 65-110. Since before and after 1982 the categories of the health status has changed, which normally induces a big jump in the whole sample estimation. Moreover, the substantial increase in life expectancy results in a significant rise in the proportion of the elderly population, which indicates the importance of the health problem when people are aging, hence the 65-110 age groups are selected.

The same procedures as in section 4.2 are applied with the subsample. I first estimate the Lee-Carter model for both the mortality and health processes (HSI with the different measures and formats). Then by projecting the future trends of mortality, a cohort life table can be created, which sequentially provides life expectancy. By including the same length projection of the HSI into a cohort life table, healthy life expectancy can be derived. Again, the simulation method to is employed analyze the uncertainty intervals by taking both process risk and parameter risk into account. All the results for the subsample are presented in appendices F, G and H. Note that Figures 58 and 59 show a smoother and more linear relation of κ_t^H ($\kappa_t^{H'}$) over time compared with κ_t^H ($\kappa_t^{H'}$) (see Figures 24 and 25) estimated in the whole sample. This implies that the selected subsample is as expected. Another interesting result shown in the subsample analysis is that they provide lower life expectancy and lower increasing rate than using the whole sample, which are shown by the first panels in tables 1 and 4, and the third and fourth rows of the first panels for females, and the third and fourth rows of the second panels for males in tables 3 and 6. This may be due to the infant death rate that has been dramatically reduced over the past century, implying that life expectancy at birth also correspondingly improved. Although statistically, the reduction in infant deaths suggests that people are living longer, if we eliminate the influence of changes in the infant death rate by examining life expectancy at age 65 or 85, it will be revealed that life expectancy over the past century has increased not as quickly as indicated by the whole sample analysis. The rest of the results for the subsample analysis are very similar to the whole sample analysis, and share the same explanations as in section 4.2. Moreover, life expectancy from both whole sample and subsample in this study are higher than life expectancy published by the NHIS in its annual technical reports, since this analysis is based on the cohort life table, whereas the NHIS estimates of life expectancy are based on a age-specific period life table.

The mean squared errors are also calculated for the subsample analysis to test the goodness-of-fit, and listed in the last rows of each panel in table 7. It can be seen that the models provide slightly higher MSEs when using the subsample, however, the differences are in very small magnitudes. Consequently, we can conclude that in the analysis for the U.S., the lengths of the mortality and health time span are not that crucial, as suggested by Lee and Carter (1992). Simularly as indicated by the mean squared error for the health estimation in the whole sample analysis, models in the subsample analysis using the level format outperforms using the logit format when defining both poor and fair health states as bad health, but not when considering only poor state as bad health. However, the right handed plot in Figure 74 for males shows that at the end of the forecasting period, healthy life expectancy forecasted using the level format of the HSI even increases above life expectancy, which contradicts the reality. This is because, when using the level data, the health of males are improved over time as suggested from the estimation, the projected $\pi_{x,t}$ possibly becomes negative at the end of the forecasting period, which results in a higher healthy life expectancy compared with life expectancy. This problem however, is solved by using the logit transformation by construction. Since transforming the

logit format back to the level format of the HSI, will always keep the HSI nonnegative, since

$$\pi_{x,t} = \frac{\exp(g_{x,t})}{1 + \exp(g_{x,t})},$$

where $g_{x,t}$ denotes $logit(\pi_{x,t})$. In this case, even the models using the logit format have slightly higher mean squared errors, they still own the advantage of keeping nonnegative projected HSI.

5 Conclusion and Discussion

This thesis incorporates the self-reported health status information from the NHIS in the United States into cohort life tables to estimate and forecast healthy life expectancy. The thesis first illustrates a so called *health status index* to measure the self-reported health status, which helps to decompose a healthy and an unhealthy part of total life expectancy. We choose the health status index from the self-reported health information instead of the disability prevalence ratio used in many of the existing literature, because it is more suitable to reflect the people's own perception about their health being relevant to the working ability. The novelty of this thesis is applying the Lee-Carter model to describe not only the mortality process but also the health process, as well as to construct its stochastic projection, corresponding to healthy life expectancy.

In this study, the health status index is measured in two different ways: only "poor" state is treated as bad health, or both "poor" and "fair" states are deemed as bad health. Moreover, the health status index is modeled by the Lee-Carter model both with its level format and its logit transformation. The empirical analysis for the United States shows that the Lee-Carter model used in the health analysis fits the data quite well, since the mean squared errors are usually very small under the two measures of the health status index with both level and logit formats. In addition, the Lee-Carter model indicates a fluctuating increasing trend of the people's health. Such increasing trends are more obvious for the elderly people. Though, such trends increase very slowly and tend to be stable.

After that, life expectancy is computed from a cohort life table, into which, by including the health status index, healthy life expectancy is estimated and projected. It is found that life expectancy and healthy life expectancy are increasing over time in the United States, which is similar to Weale and Khoman (2006), who examined healthy life expectancy for the United Kingdom. However, life expectancy grows faster than healthy life expectancy in the U.K. which is opposite to the results found for the United States in this thesis. Although, much literature about the U.S. argues that the proportion of the disabled increases in the population, which induces the health care expenditure growing dramatically and lowering the increase in healthy life expectancy, see, for instance, Robine and Ritchie (1991), Zweifel, Felder, and Meiers (1999) and Stearns and Norton (2004). As this thesis argued, healthy life expectancy estimated and forecasted from this study is not based on the disability prevalence, but on the self-reported health status, which is more relevant to people's own perception of ability to work. A higher increasing rate of healthy life expectancy than life expectancy obtained in this thesis is in line with the results provided by Manton, Stallard, and Tolley (1991), and also consistent with the health estimates by Duggan and Imberman (2006), who found the health condition in the U.S. is increasing on average, which in turn yields a higher increasing rate of healthy life expectancy.

Another novelty of this thesis is the inclusion of uncertainty intervals for life expectancy and healthy life expectancy. The analysis on considering both process and parameter risks are proceeded by a simulation technique. We find that the expected simulated healthy life expectancy increases faster than the expected simulated life expectancy with larger uncertainty.

Moreover, males' expected simulated life expectancy and healthy life expectancy are lower than females', but increase faster over time with larger uncertainty. Note that in this study, simulation risks are negligible due to the small simulation errors.

It is worth mentioning that either using the level format or the logit format for the health status index fits the data quite well and they do not provide significant different estimates and forecasts of healthy life expectancy as expected. Healthy life expectancy obtained using the logit health status index are slightly lower, increasing slower and owning narrower confidence intervals than using the level format. Moreover, the logit transformed model is superior to the level format model by construction, since it always provides a nonnegative projected health status index, which is desirable in any practical analysis.

So far, this thesis provides evidence that the Lee-Carter model fits the data well when modeling the health and mortality processes in a stochastic way, which can yield reasonable life expectancy from a cohort life table and the corresponding healthy life expectancy. Notwithstanding, there are several interesting extensions that can be made in future research. First, it is possible to decompose total life expectancy, not only into a healthy and an unhealthy part, but into multiple health states of interest using the self-reported health information. This can give a more detailed impression on the health effects in people's residual lifetime. However, the self-reported health information is still a controversial data set being used, which need to be carefully examined in the future. Second, I model the health status index by the Lee-Carter model and derive the stochastic projection by classic ARIMA (0,1,0) process for the latent variable, which is possibly more suitable to be measured by other time series models. Moreover, except parameter risk caused by the ARIMA model, which is examined in this study, parameter risk caused by the Lee-Carter model (21) are not yet included in the current study. Another aspect which is worth to be investigated in the future is the joint effects of both females and males.

References

BOOTH, H., J. MAINDONALD, AND L. SMITH (2002): "Applying Lee-Carter under Conditions of Variable Mortality Decline," .

CAIRNS, A. J. G. (2007): "A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States," *Heriot-Watt University, and Pensions Institute Discussion Paper*, (PI-0701), Working Paper.

CAIRNS, A. J. G., D. BLAKE, AND K. DOWD (2006): "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration," *Journal of Risk and Insurance*, 73(4), 687–718.

CHIANG, C. L. (1984): *The Life Table and its Applications*. R.E. Krieger Pub. Co., Malabar, Fla.

COSTA, D. L. (2002): "Changing Chronic Disease Rates and Long-term Declines in Functional Limitation Among Older Men," *Demography*, 64(1), 119–137.

CURRIE, I. D., M. DURBAN, AND P. H. C. EILERS (2004): "Smoothing and Forecasting Mortality Rates," *Statistical Modeling*, 4, 279–298.

DENUIT, M., AND A. GODERNIAUX (2005): "Closing and Projecting Life Tables using Log-linear Models," *Bulletin of the Swiss Association of Actuaries*, 1(4), 29–49.

DUGGAN, M., AND S. A. IMBERMAN (2006): *Health in Older Ages: The Causes and Consequences of Declining Disability among the Elderly*forthcoming Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity. University of Chicago Press, Forthcoming.

GMEZ, P. G., AND A. L. NICOLS (2006): "Health Shocks, Employment and Income in the Spanish Labor Markets," *Health Economics*, 15, 997–1000.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

HASTIE, T., AND R. TIBSHIRANI (1986): "Generalized Additive Models," *Statistical Science*, 1, 297–310.

IMAI, K., AND S. SONEJI (2007): "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension," *Journal of the American Statistical Association*, 102, 1199–1211.

KATZ, S., L. BRANCH, M. BRANSON, J. PAPSIDERO, J. BECK, AND D. GREER (1983): "Active Life Expectancy," *New England Journal of Medicine*, 309(20), 1218–1224.

KIRKBY, J. G., AND I. D. CURRIE (2010): "Smooth Models of Mortality with Period Shocks," *Statistical Modeling*, 10(2), 177196.

LAKDAWALLA, D., D. P. GOLDMAN, AND J. BHATTACHARYA (2004): "Are The Young Becoming More Disabled?," *Health Affairs*, 23, 68–76.

LECHNER, M., AND R. VAZQUEZ-ALVAREZ (2003): "The Effect of Disability on Labour Market Outcomes in Germany: Evidence from Matching," IZA Discussion Papers 967, Institute for the Study of Labor (IZA).

LEE, R. D., AND L. R. CARTER (1992): "Modeling and Forecasting U.S. Mortality," *Journal of the American Statistical Association*, 87, 659671.

LEE, R. D., AND T. MILLER (2001): "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality," *Demography*, 38(4), 537–549.

LIVRE, A., N. BROUARD, AND C. HEATHCOTE (2003): "The Estimation of Health Expectancies from Cross-longitudinal Surveys," *Mathematical Population Studies*, 10, 211–248.

MANTON, K. G., L. CORDER, AND E. STALLARD (1993): "Changes in the Use of Personal Assistance and Special Equipment from 1982 to 1989: Results from the 1982 and 1989 NLTCS," *Gerontologist*, 33(2), 168–76.

MANTON, K. G., E. S. CORNELIUS, AND M. A. WOODBURY (1995): "Nursing Home Residents: A Multivariate Analysis of Their Medical, Behavioral, Psychosocial, and Service use Characteristics," *Journal of Gerontology*, 50(5), 242–251.

MANTON, K. G., X. GU, AND V. L. LAMB (2006): "Long-Term Trends in Life Expectancy and Active Life Expectancy in the United States," *Population and Development Review*, 32(1), 81–105.

MANTON, K. G., X. GU, AND G. R. LOWRIMORE (2008): "Cohort Changes in Active Life Expectancy in the US Elderly Population: Experience From the 1982-2004 National Long Term Care Survey," *Journal of Gerontology*, 63B(5), 269–281.

MANTON, K. G., AND K. C. LAND (2000): "Active Life Expectancy Estimates for the U.S. Elderly Population: A Multidimensional Continuous-Mixture Model of Functional Change Applied to Completed Cohorts, 1982-1996," *Demography*, 37(3), 253–265.

MANTON, K. G., AND E. STALLARD (1991): "Cross-sectional Estimates of Active Life Expectancy for the U.S. Elderly and Oldest-Old Populations," *Journal of gerontology*, 46(3), 170–182.

MANTON, K. G., E. STALLARD, AND L. CORDER (1997): "Changes in the Age Dependence of Mortality and Disability: Cohort and Other Determinants," *Demography*, 34(1), 135–157.

MANTON, K. G., E. STALLARD, AND D. H. TOLLEY (1991): "Limits to Human Life Expectancy: Evidence, Prospects, and Implications," *Population and Development Review*, 17(4), 603–637.

MANTON, K. G., AND M. A. WOODBURY (1982): "A New Procedure for Analysis of Medical Classification," *Methods of Information in Medicine*, 21, 183–244.

MATHERS, C. D., AND J. ROBINE (1997): "How Good is Sullivan's Method for Monitoring Changes in Population Health Expectancies?," *Journal of Epidemiology and Community Health*, 51(1), 80–86.

MOLLA, M. T., D. K. WAGENER, AND J. H. MADANS (2001): "Summary Measures of Population Health: Methods for Calculating Healthy Life Expectancy," *Healthy People 2010 Statistical Notes*, (21), 1–11.

PITACCO, E., M. DENUIT, S. HABERMAN, AND A. OLIVIERI (2009): *Modeling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press.

PORTRAIT, F., M. LINDEBOOM, AND D. DEEG (2001): "Life Expectancies in Specific Health States Results from a Joint Model of Health Status and Mortality of Older Persons," *Demography*, 38(4), 525–536.

PRESTON, S. H., P. HEUVELINE, AND M. GUILLOT (2001): *Demography: Measuring and Modeling Population Processes*. Blackwell Publishers, London.

RENSHAW, A. E., AND S. HABERMAN (2003a): "Lee-Carter Mortality Forecasting: A Parallel Generalized Linear Modeling Approach for England and Wales Mortality Projections," *Applied Statistics*, 52(1), 119–137.

——— (2003b): "Lee-Carter Mortality Forecasting with Age-specific Enhancement," *Insurance: Mathematics and Economics*, 33(2), 255–272.

——— (2003c): "On the Forecasting of Mortality Reduction Factors," *Insurance: Mathematics and Economics*, 32, 379401.

——— (2005): "A Cohort-based Extension to the Lee-Carter Model for Mortality Reduction Factors," *Insurance: Mathematics and Economics*, 38(3), 557–570.

ROBINE, J. M., AND K. RITCHIE (1991): "Healthy Life Expectancy: Evaluation of A New Global Indicator for Change in Population Health," *British Medical Journal*, 302, 457–460.

ROGERS, A., R. G. ROGERS, AND A. BELANGER (1990): "Longer Life but Worse Health? Measurement and Dynamics," *The Gerontologist*, 30(5), 640–649.

STEARNS, S. C., AND E. C. NORTON (2004): "Time to Include Time to Death? The Future of Health Care Expenditure Predictions," *Health Economics*, 13, 315327.

SULLIVAN, D. F. (1971): "A Single Index of Mortality and Morbidity," *HSMHA Health Reports*, 86, 347–354.

TULJAPURKAR, S., N. LI, AND C. BOE (2000): "A Universal Pattern of Mortality Decline in the G7 Countries," *Nature*, 405, 789792.

WEALE, MARTIN, R., AND E. KHOMAN (2006): "Incidence-based Estimates of Healthy Life Expectancy for the United Kingdom: Coherence between Transition Probabilities and Aggregate Life Tables," Niesr discussion papers, National Institute of Economic and Social Research.

WILMOTH, J. R. (1993): "Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change," *U.C. Berkeley Technical Report*.

WOODBURY, M. A., AND K. G. MANTON (1989): "Grade of Membership Analysis of Depression-Related Psychiatric Disorders," *Sociological Methods and Research*, 18(1), 126–163.

ZWEIFEL, P., S. FELDER, AND M. MEIERS (1999): "Ageing of Population and Health Care Expenditure: A Red Herring?," *Health Economics*, 8, 485–496.

A Tables

A.1 For the Whole Sample

Table 1: Point estimates and forecasts of life expectancy and healthy life expectancy at particular ages and years using the whole sample.
'IR' denotes the Increasing Rate of the LE or HLE

	Female						Male					
Ages	1985	1995	2005	2015	2025	IR %	1985	1995	2005	2015	2025	IR%
	Life Expectancy											
0	84.10	85.36	86.48	87.48	88.39	5.11	80.59	82.33	83.87	85.24	86.46	7.28
25	58.18	59.21	60.19	61.11	61.98	6.53	54.01	55.59	57.05	58.41	59.66	10.46
45	37.21	38.23	39.21	40.15	41.04	10.29	32.86	34.47	35.99	37.42	38.74	17.91
65	18.94	19.69	20.43	21.16	21.86	15.42	15.27	16.42	17.53	18.61	19.63	28.57
85	6.19	6.51	6.82	7.13	7.43	20.17	5.05	5.34	5.64	5.93	6.22	23.19
	Healthy Life Expectancy (Level & Poor)											
0	82.01	83.49	84.84	86.09	87.26	6.40	79.34	81.43	83.35	85.12	86.76	9.35
25	55.82	57.00	58.13	59.23	60.29	8.01	52.24	54.06	55.80	57.47	59.05	13.04
45	34.94	36.06	37.16	38.22	39.26	12.37	30.95	32.73	34.45	36.11	37.70	21.82
65	17.39	18.19	18.98	19.78	20.56	18.24	13.94	15.15	16.34	17.52	18.66	33.87
85	5.46	5.78	6.11	6.43	6.75	23.58	4.44	4.74	5.05	5.37	5.68	28.09
	Healthy Life Expectancy (Logit & Poor)											
0	81.69	83.07	84.31	85.42	86.43	5.80	78.92	80.82	82.50	84.00	85.33	8.11
25	55.62	56.75	57.83	58.85	59.82	7.54	52.09	53.81	55.41	56.90	58.28	11.88
45	34.79	35.88	36.94	37.95	38.91	11.85	30.88	32.62	34.27	35.81	37.25	20.63
65	17.30	18.07	18.84	19.59	20.33	17.53	13.90	15.10	16.28	17.41	18.49	32.97
85	5.45	5.76	6.07	6.38	6.69	22.81	4.43	4.73	5.04	5.35	5.65	27.71
	Healthy Life Expectancy (Level & Fair)											
0	73.87	75.66	77.34	78.92	80.43	8.87	72.50	74.65	76.65	78.52	80.28	10.74
25	47.80	49.19	50.54	51.85	53.12	11.13	45.92	47.72	49.46	51.12	52.72	14.80
45	28.08	29.23	30.35	31.43	32.49	15.68	25.60	27.26	28.88	30.44	31.94	24.77
65	13.02	13.68	14.31	14.94	15.54	19.30	10.58	11.58	12.57	13.53	14.46	36.65
85	4.03	4.17	4.30	4.43	4.54	12.68	3.28	3.44	3.59	3.73	3.87	18.10
	Healthy Life Expectancy (Logit & Fair)											
0	73.23	74.70	75.98	77.11	78.11	6.66	71.53	73.37	75.00	76.45	77.73	8.67
25	47.62	48.82	49.94	50.97	51.93	9.05	45.48	47.09	48.59	49.97	51.25	12.68
45	28.08	29.12	30.12	31.05	31.93	13.73	25.41	26.96	28.42	29.80	31.07	22.28
65	13.06	13.67	14.26	14.82	15.36	17.62	10.54	11.49	12.41	13.29	14.13	33.97
85	4.03	4.18	4.31	4.43	4.54	12.72	3.29	3.45	3.60	3.74	3.87	17.63

Table 2: Simulation errors for the whole sample estimation with 2000 times simulations

	Female (%)				
	LE	HLE (Level&Poor)	HLE (Logit&Poor)	HLE (Level&Fair)	HLE (Logit&Fair)
Process Risk	0.01	0.06	0.01	0.01	0.01
Process&Parameter Risk	0.04	0.18	0.05	0.03	0.02
	Male (%)				
	LE	HLE (Level&Poor)	HLE (Logit&Poor)	HLE (Level&Fair)	HLE (Logit&Fair)
Process Risk	0.03	0.08	0.04	0.04	0.02
Process&Parameter Risk	0.09	0.22	0.14	0.13	0.06

Table 3: Simulated life expectancy and healthy life expectancy at age 65 and particular years when considering different risks and using the whole sample.
'IR' denotes the Increasing Rate; Notations including "(PP)" denotes the simulation includes both the process risk and the parameter risk, otherwise just includes the process risk of e_t.

	Female					
65	1985	1995	2005	2015	2025	IR %
SimLE	18.95	19.70	20.44	21.16	21.87	15.41
SimLE(PP)	18.94	19.68	20.42	21.14	21.84	15.31
SimHLE (Level & Poor)	17.38	18.18	18.97	19.77	20.57	18.35
SimHLE(PP) (Level &Poor)	17.36	18.16	18.94	19.73	20.49	18.03
SimHLE (Logit & Poor)	17.22	17.98	18.73	19.47	20.20	17.31
SimHLE(PP) (Logit & Poor)	17.20	17.94	18.68	19.40	20.09	16.77
SimHLE(Level & Fair)	13.04	13.68	14.31	14.92	15.53	19.12
SimHLE(PP)(Level & Fair)	13.02	13.67	14.32	14.94	15.53	19.30
SimHLE (Logit & Fair)	13.03	13.63	14.21	14.76	15.29	17.31
SimHLE(PP) (Logit & Fair)	13.03	13.62	14.18	14.73	15.23	16.93
	Male					
65	1985	1995	2005	2015	2025	IR %
SimLE	15.25	16.39	17.50	18.58	19.60	28.54
SimLE(PP)	15.29	16.45	17.56	18.63	19.65	28.48
SimHLE (Level & Poor)	13.92	15.13	16.32	17.49	18.63	33.77
SimHLE(PP) (Level &Poor)	13.95	15.14	16.32	17.46	18.59	33.22
SimHLE (Logit & Poor)	13.80	14.97	16.11	17.21	18.26	32.30
SimHLE(PP) (Logit & Poor)	13.72	14.81	15.87	16.89	17.84	30.04
SimHLE(Level & Fair)	10.57	11.56	12.55	13.51	14.44	36.69
SimHLE(PP)(Level & Fair)	10.59	11.58	12.56	13.51	14.45	36.45
SimHLE (Logit & Fair)	10.51	11.44	12.35	13.21	14.03	33.49
SimHLE(PP) (Logit & Fair)	10.49	11.41	12.30	13.13	13.90	32.50

A.2 For the Subsample

Table 4: Point estimates and forecasts of life expectancy and healthy life expectancy at particular ages and years using the subsample (Age 65-110, Period 1982-2006)
'IR' denotes the Increasing Rate of LE or HLE.

	Female						Male					
Ages	1985	1995	2005	2015	2025	IR %	1985	1995	2005	2015	2025	IR%
	Life Expectancy											
65	18.81	19.25	19.68	20.11	20.52	9.07	15.10	16.24	17.35	18.43	19.47	28.91
75	11.73	12.03	12.33	12.63	12.92	10.14	9.06	9.77	10.48	11.18	11.87	31.06
85	6.32	6.46	6.61	6.75	6.88	8.90	4.96	5.27	5.59	5.90	6.22	25.43
	Healthy Life Expectancy (Level & Poor)											
65	17.07	17.84	18.63	19.43	20.24	18.61	13.57	15.01	16.48	17.97	19.47	43.49
75	10.32	10.86	11.40	11.96	12.53	21.41	7.92	8.76	9.64	10.54	11.46	44.78
85	5.41	5.66	5.91	6.16	6.41	18.39	4.24	4.60	4.97	5.35	5.73	35.13
	Healthy Life Expectancy (Logit & Poor)											
65	17.10	17.83	18.50	19.13	19.71	15.30	13.66	15.09	16.41	17.65	18.81	37.68
75	10.35	10.87	11.36	11.82	12.24	18.31	7.97	8.83	9.67	10.47	11.25	41.09
85	5.42	5.67	5.90	6.12	6.33	16.75	4.24	4.62	4.99	5.36	5.72	34.74
	Healthy Life Expectancy (Level & Fair)											
65	12.88	13.87	14.87	15.89	16.93	31.48	10.14	11.51	12.91	14.32	15.73	55.15
75	7.56	8.09	8.63	9.18	9.75	28.93	5.77	6.44	7.11	7.79	8.47	46.68
85	3.95	4.15	4.35	4.55	4.76	20.62	3.19	3.35	3.50	3.65	3.79	18.94
	Healthy Life Expectancy (Logit & Fair)											
65	12.89	13.85	14.77	15.64	16.46	27.74	10.18	11.55	12.86	14.09	15.23	49.57
75	7.57	8.09	8.61	9.11	9.61	26.89	5.79	6.46	7.11	7.75	8.36	44.30
85	3.95	4.15	4.35	4.55	4.74	20.07	3.19	3.35	3.50	3.64	3.77	18.09

Table 5: Simulation errors for the subsample estimation with 3000 simulations

	Female (%)				
	LE	HLE (Level&Poor)	HLE (Logit&Poor)	HLE (Level&Fair)	HLE (Logit&Fair)
Process Risk	0.05	0.07	0.03	0.08	0.04
Process&Parameter Risk	0.15	0.23	0.15	0.24	0.15
	Male (%)				
	LE	HLE (Level&Poor)	HLE (Logit&Poor)	HLE (Level&Fair)	HLE (Logit&Fair)
Process Risk	0.23	0.20	0.14	0.15	0.10
Process&Parameter Risk	0.71	0.62	0.48	0.46	0.31

Table 6: Simulated life expectancy and healthy life expectancy at age 65 and particular years when considering different risks and using the subsample (Age 65-110, Period 1982-2006)
'IR' denotes the Increasing Rate; Notations including "(PP)" denotes the simulation includes both the process risk and the parameter risk, otherwise just includes the process risk of e_t.

	Female					
65	1985	1995	2005	2015	2025	IR %
SimLE	18.81	19.24	19.67	20.09	20.50	8.98
SimLE(PP)	18.80	19.22	19.63	20.03	20.42	8.65
SimHLE (Level & Poor)	17.07	17.85	18.64	19.43	20.22	18.49
SimHLE(PP) (Level & Poor)	17.05	17.81	18.58	19.36	20.14	18.15
SimHLE (Logit & Poor)	17.01	17.71	18.36	18.97	19.54	14.89
SimHLE(PP) (Logit & Poor)	16.95	17.59	18.16	18.68	19.16	13.02
SimHLE(Level & Fair)	12.87	13.87	14.86	15.87	16.90	31.29
SimHLE(PP)(Level & Fair)	12.86	13.82	14.81	15.79	16.79	30.57
SimHLE (Logit & Fair)	12.86	13.81	14.71	15.57	16.38	27.41
SimHLE(PP) (Logit & Fair)	12.84	13.76	14.62	15.41	16.14	25.67
	Male					
65	1985	1995	2005	2015	2025	IR %
SimLE	15.06	16.17	17.26	18.28	19.27	27.94
SimLE(PP)	15.15	16.27	17.34	18.35	19.28	27.25
SimHLE (Level & Poor)	13.56	14.99	16.44	17.91	19.39	43.01
SimHLE(PP) (Level & Poor)	13.51	14.90	16.29	17.69	19.06	41.05
SimHLE (Logit & Poor)	13.53	14.87	16.14	17.34	18.48	36.59
SimHLE(PP) (Logit & Poor)	13.56	14.85	16.01	17.08	18.03	32.94
SimHLE(Level & Fair)	10.16	11.52	12.88	14.26	15.64	53.98
SimHLE(PP)(Level & Fair)	10.16	11.52	12.89	14.24	15.57	53.28
SimHLE (Logit & Fair)	10.12	11.41	12.66	13.84	14.92	47.46
SimHLE(PP) (Logit & Fair)	10.12	11.41	12.59	13.65	14.58	44.00

Table 7: Mean Squared Errors (MSE) of Different models in the Health analysis

	Mortality Estimates			
	Female (%)		Male(%)	
Whole Sample	0.02		0.02	
Subsample	0.04		0.04	
	Health Estimates			
	Level, Poor	Logit, Poor	Level, Fair	Logit, Fair
	Female (%)			
Whole Sample	0.01	0.01	0.04	0.32
SubSample	0.03	0.03	0.07	0.99
	Male(%)			
Whole Sample	0.02	0.02	0.05	0.32
SubSample	0.05	0.05	0.11	1.05

B Data Description

B.1 Mortality Description

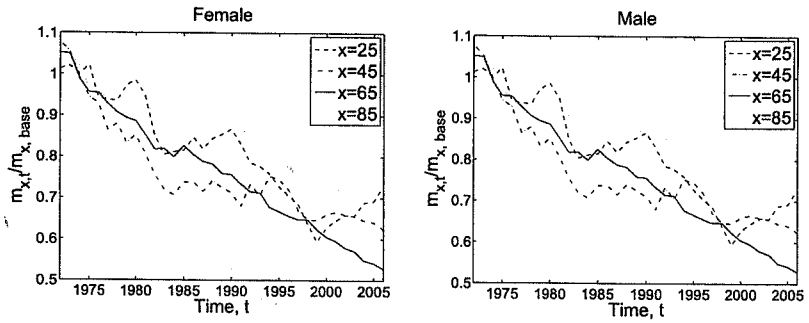

Figure 1: Raw Mortality Rates for U.S. Relative to 1972-1976 Average Mortality Rate at certain ages. '- - -': $x = 25$; '-.-': $x = 45$; '—': $x = 65$; '...': $x = 85$.

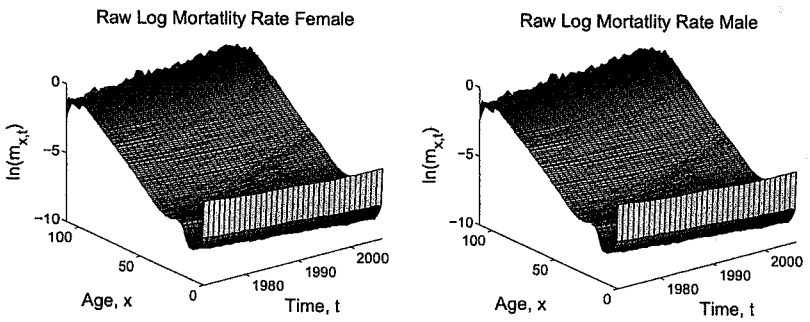

Figure 2: Log Raw Central Death Rate

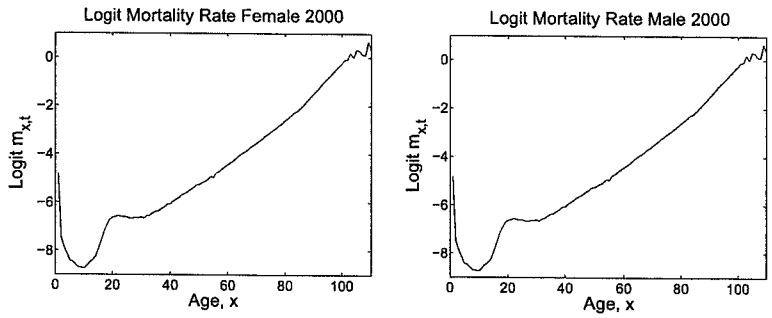

Figure 3: Logit Central Death Rate in Year 2000

B.2 Health Description

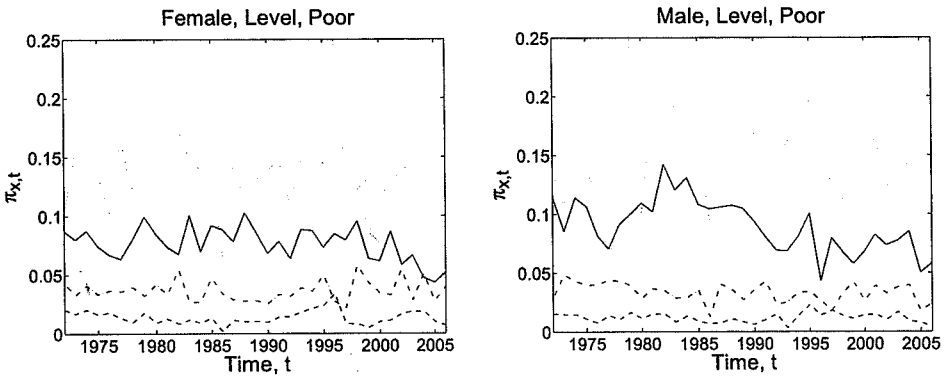

Figure 4: Raw Health Status Index (Cutting at Poor) at certain ages. '- - -': $x = 25$; '-.-': $x = 45$; '—': $x = 65$; '...': $x = 85$.

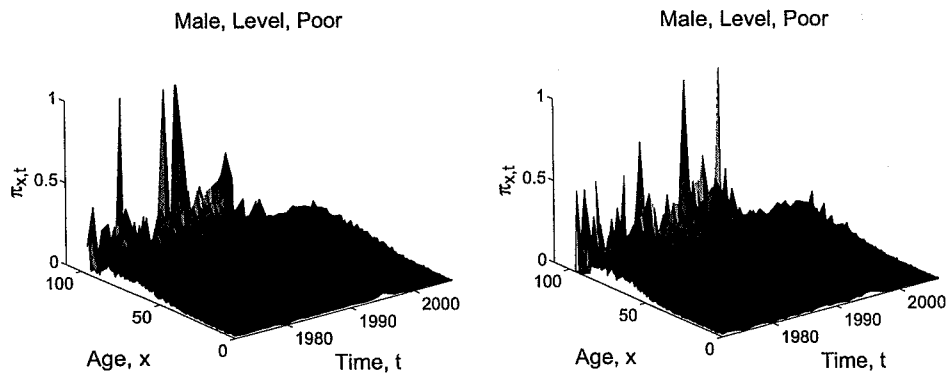

Figure 5: Raw Health Status Index (Cutting at Poor)

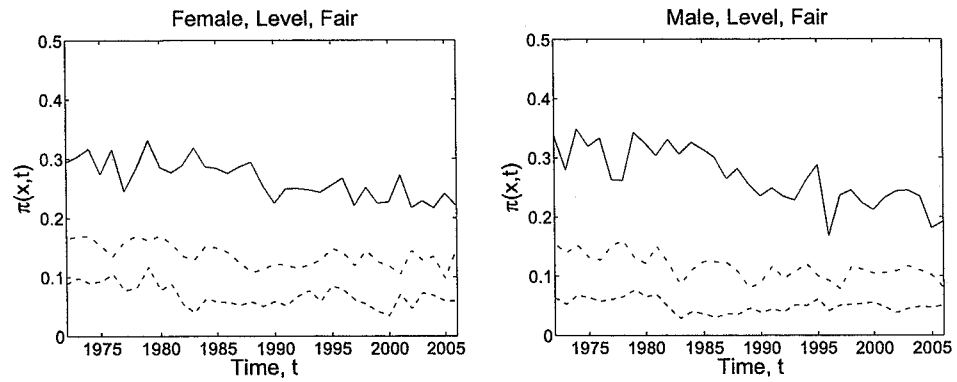

Figure 6: Raw Health Status Index (Cutting at Fair) at certain ages. '- - -': $x = 25$; '-.-': $x = 45$; '—': $x = 65$; '...': $x = 85$.

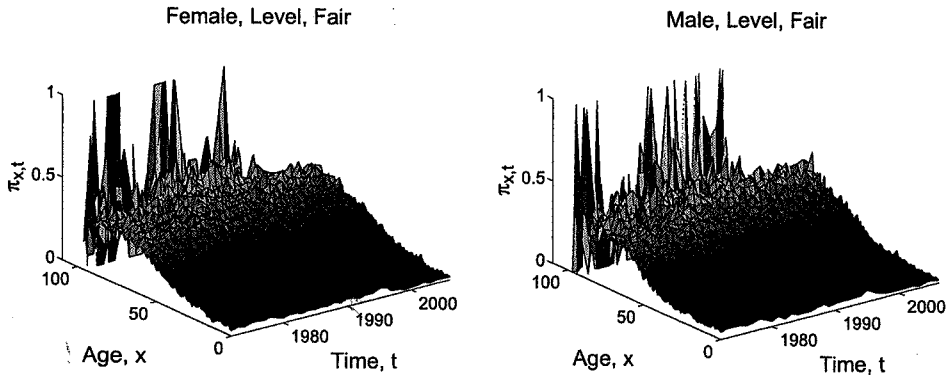

Figure 7: Raw Health Status Index (Cutting at Fair).

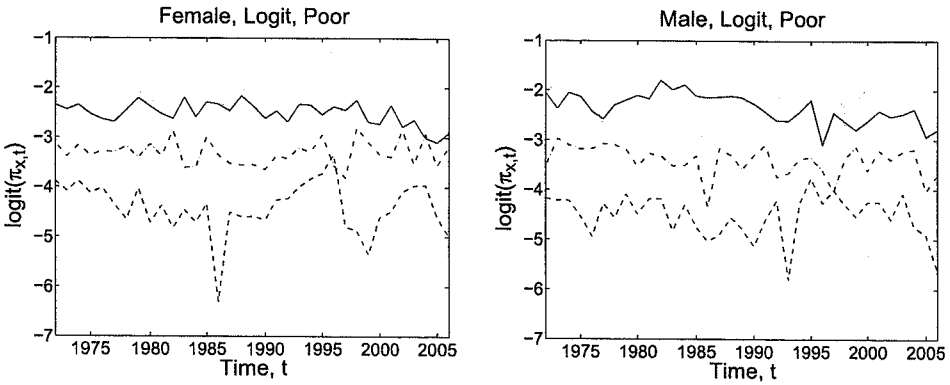

Figure 8: Logit Raw Health Status Index (Cutting at Poor) at certain ages. '- - -': $x = 25$; '-.-': $x = 45$; '—': $x = 65$; '...': $x = 85$

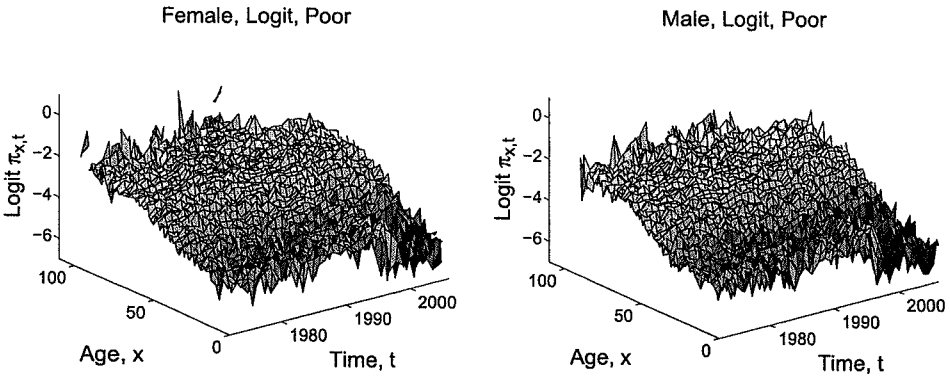

Figure 9: Logit Health Status Index (Cutting at Poor)

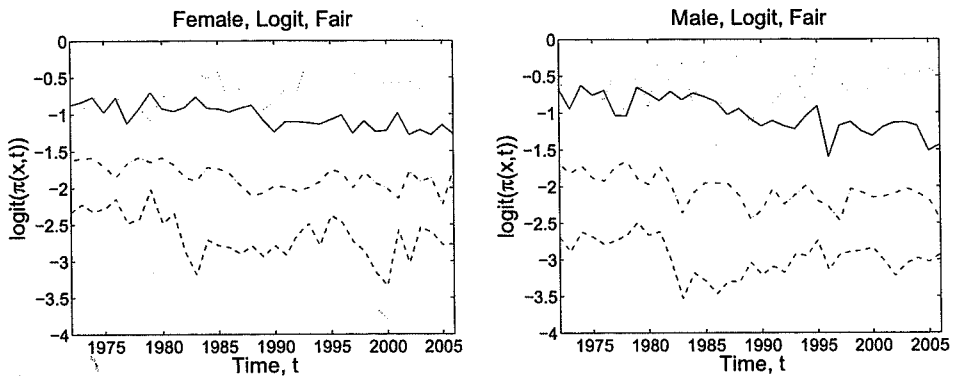

Figure 10: Logit Raw Health Status Index (Cutting at Fair) at certain ages. '- - -': $x = 25$; '-.-': $x = 45$; '—': $x = 65$; '...': $x = 85$.

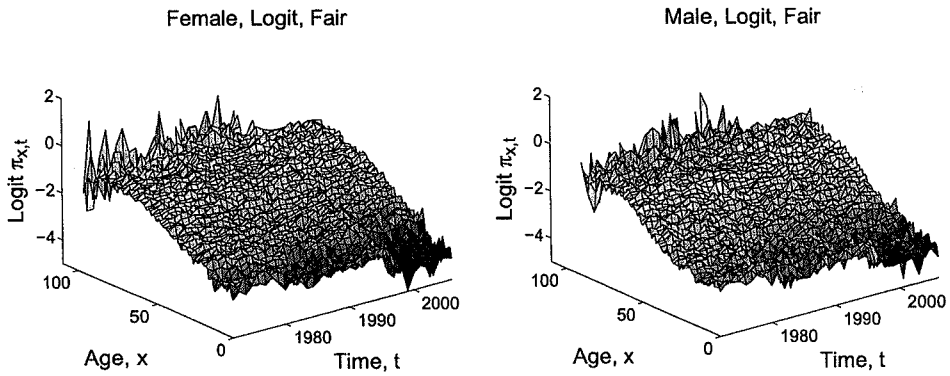

Figure 11: Logit Health Status Index (Cutting at Fair)

C Mortality Estimates and Forecasts using the Whole Sample

C.1 Estimates

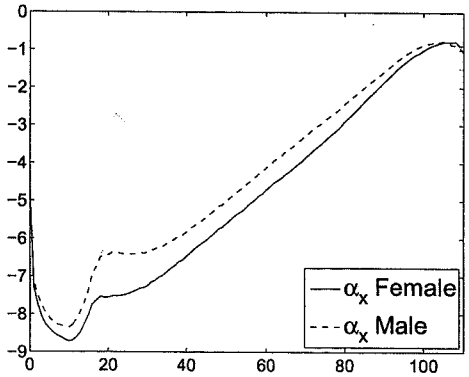

Figure 12: Estimated $\hat{\alpha}_x$ for Mortality Rate

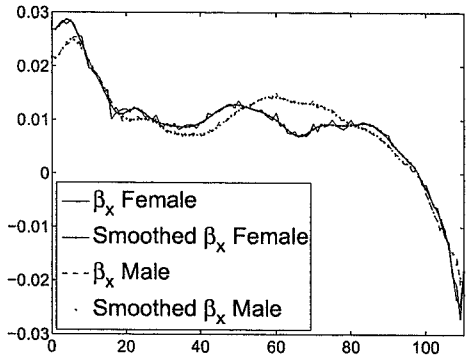

Figure 13: Estimated and Smoothed $\hat{\beta}_x$ for Mortality Rate

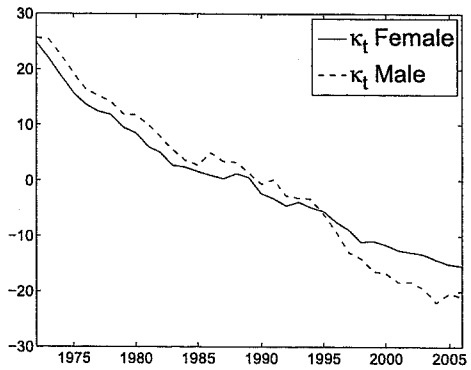

Figure 14: Estimated $\hat{\kappa}_x$ for Mortality Rate

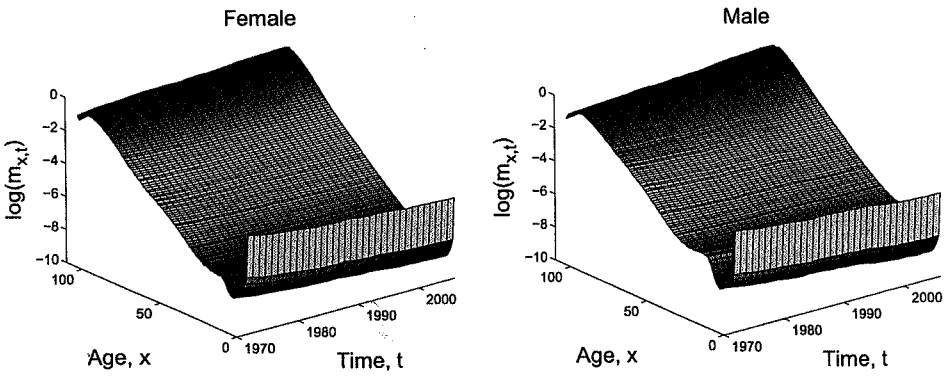

Figure 15: Estimated Log Mortality Rate, $log(m_{x,t})$

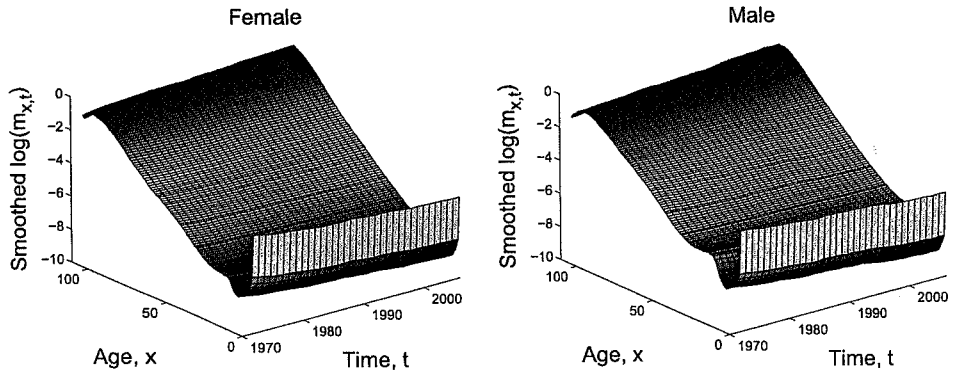

Figure 16: Smoothed Log Mortality Rate, $log(m_{x,t})$

C.2 Forecasts

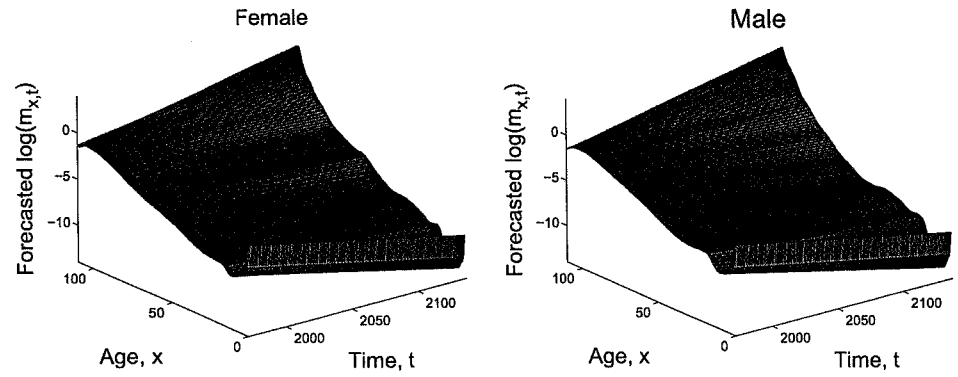

Figure 17: Forecasted Log Mortality Rate, $log(m_{x,t})$

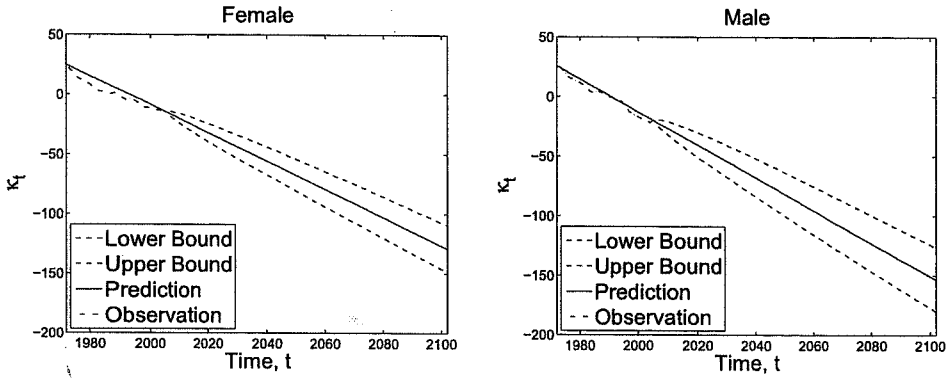

Figure 18: Forecasted κ_x and its 95% Confidence Interval

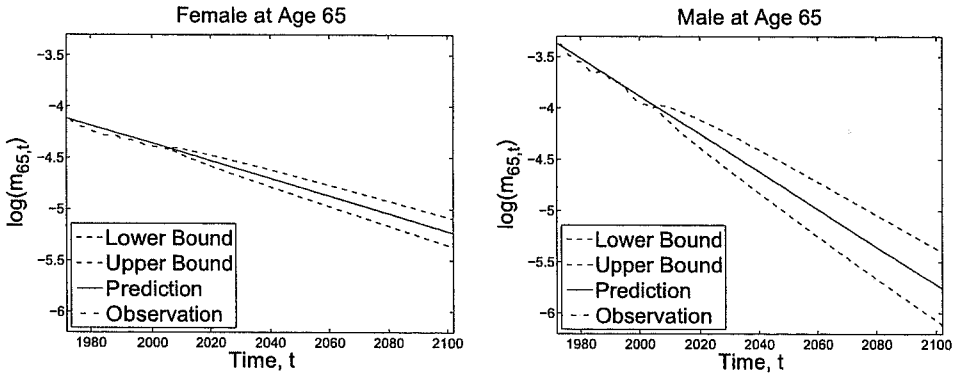

Figure 19: Forecasted Log Mortality Rate at Age 65 ($log(m_{65,t})$) and its 95% Confidence Interval

D Health Estimates and Forecasts using the Whole Sample

D.1 Estimates

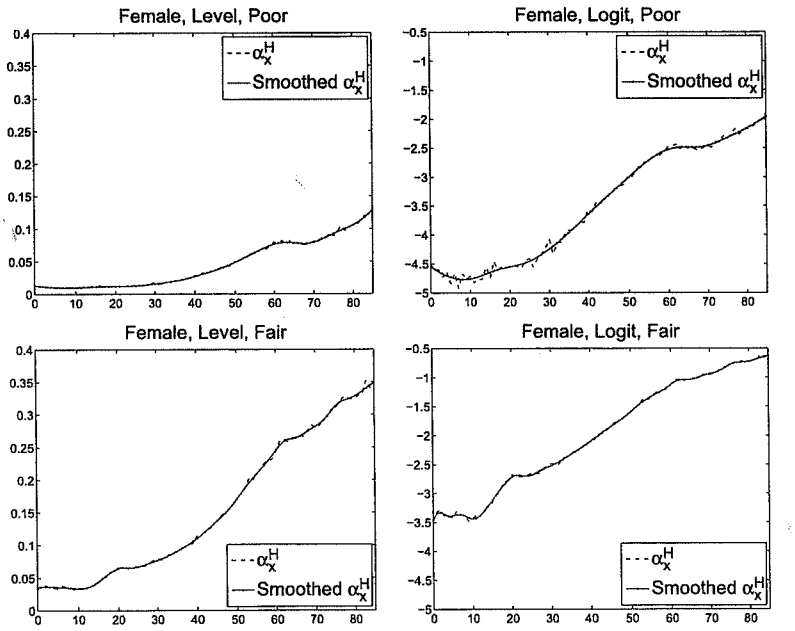

Figure 20: Estimated and Smoothed α_x^H for Female Health Status Index

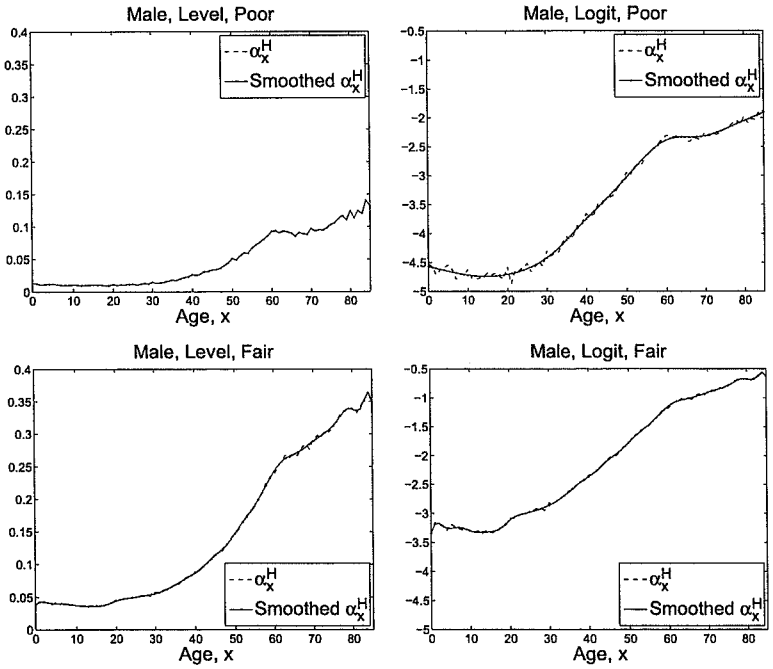

Figure 21: Estimated and Smoothed α_x^H for Male Health Status Index

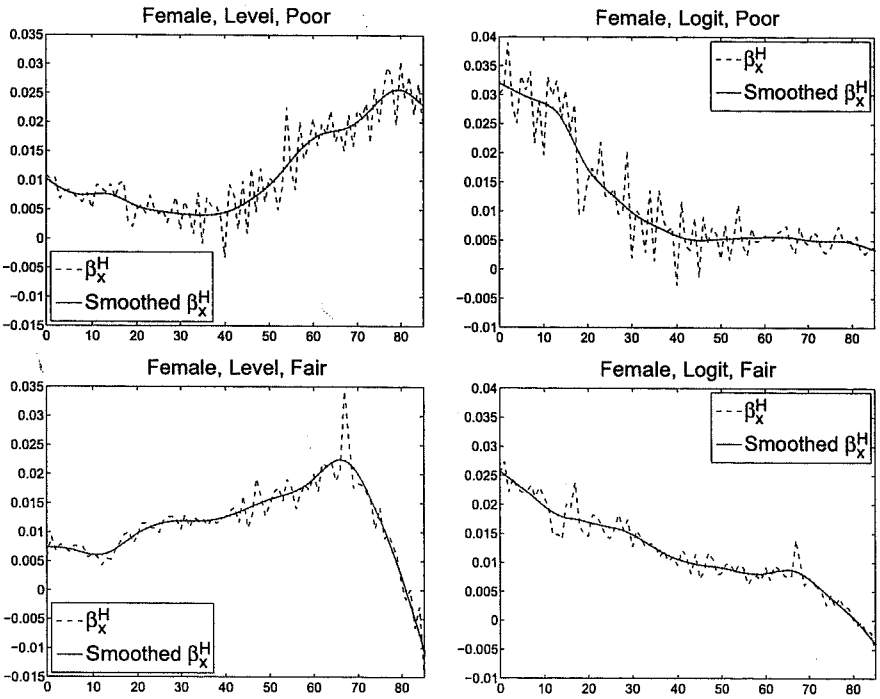

Figure 22: Estimated and Smoothed β_x^H for Female Health Status Index

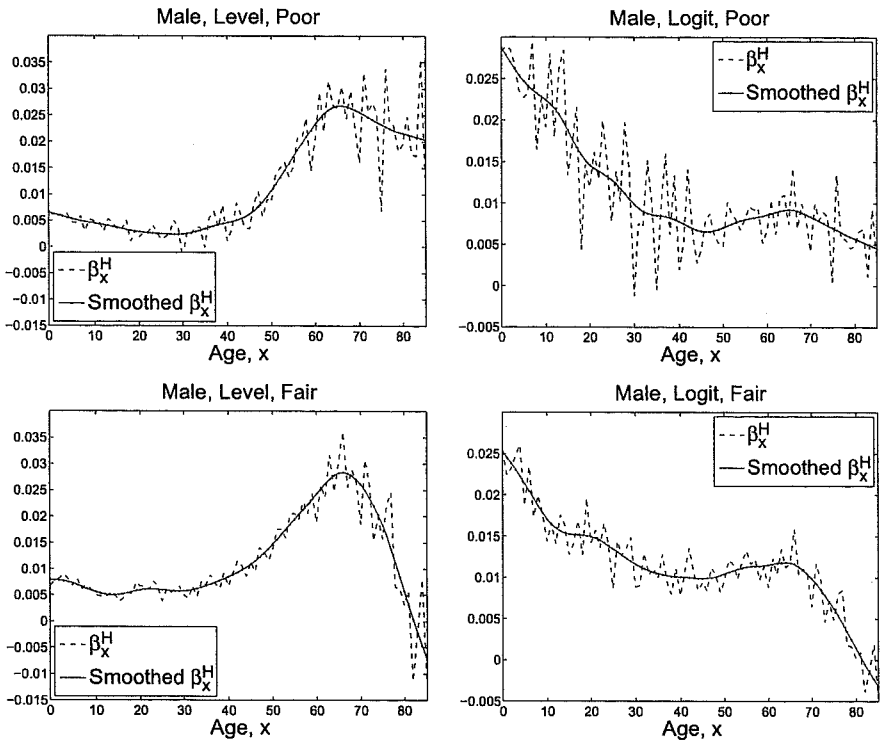

Figure 23: Estimated and Smoothed β_x^H for Male Health Status Index

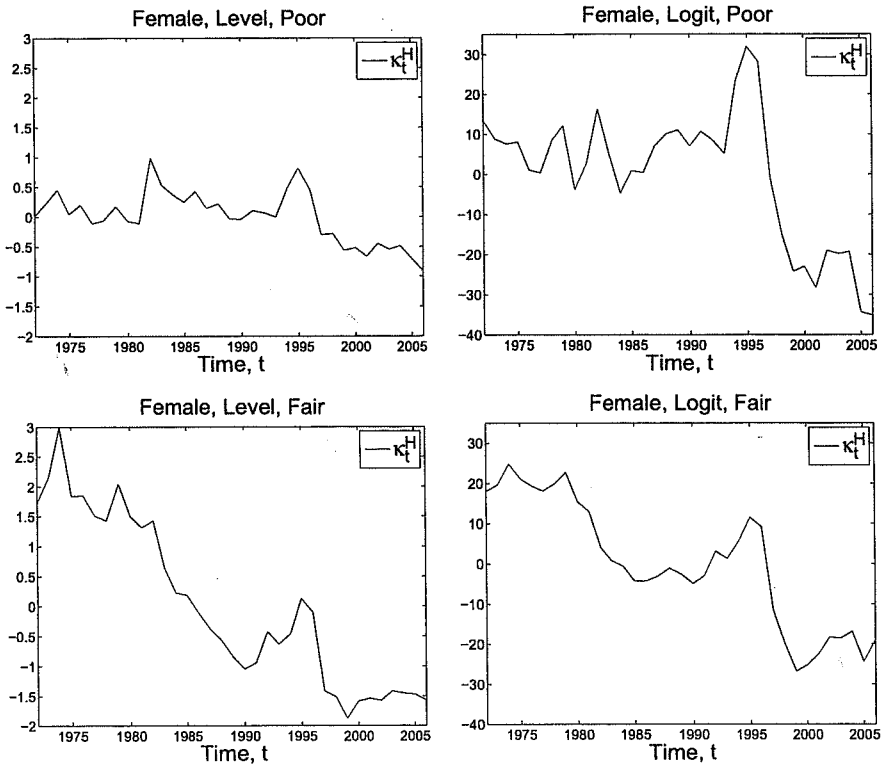

Figure 24: Estimated κ_x^H for Female Health Status Index

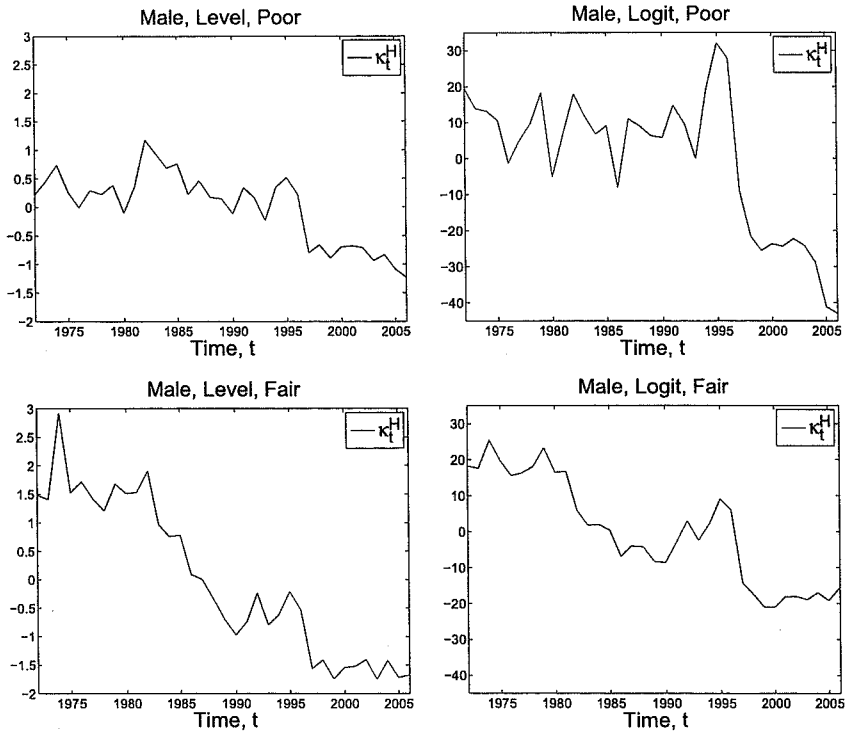

Figure 25: Estimated κ_x^H for Male Health Status Index

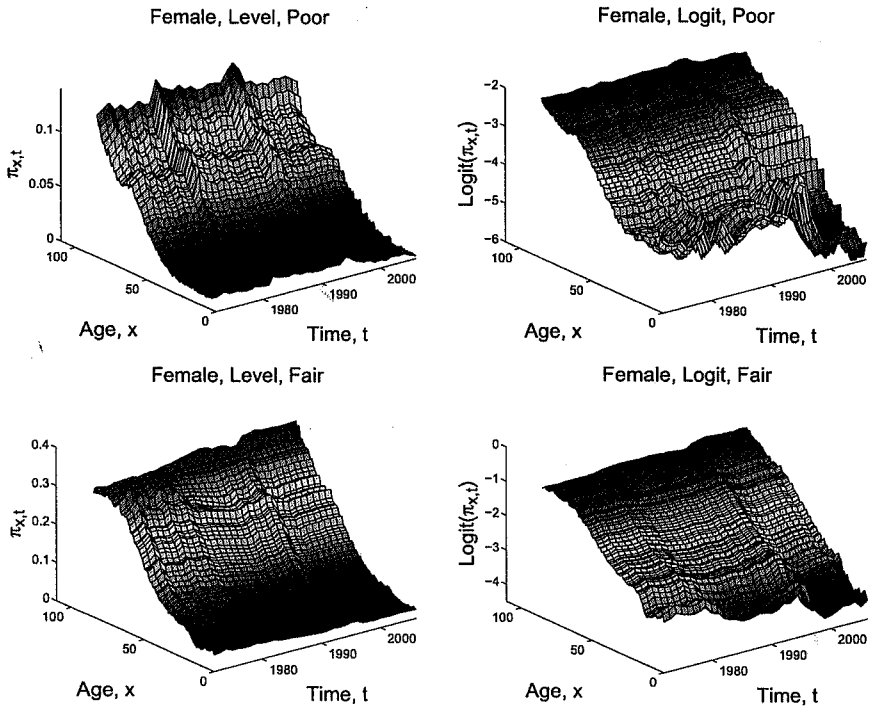

Figure 26: Estimated Health Status Index π_x for Females both in the Level and the Logit Formats

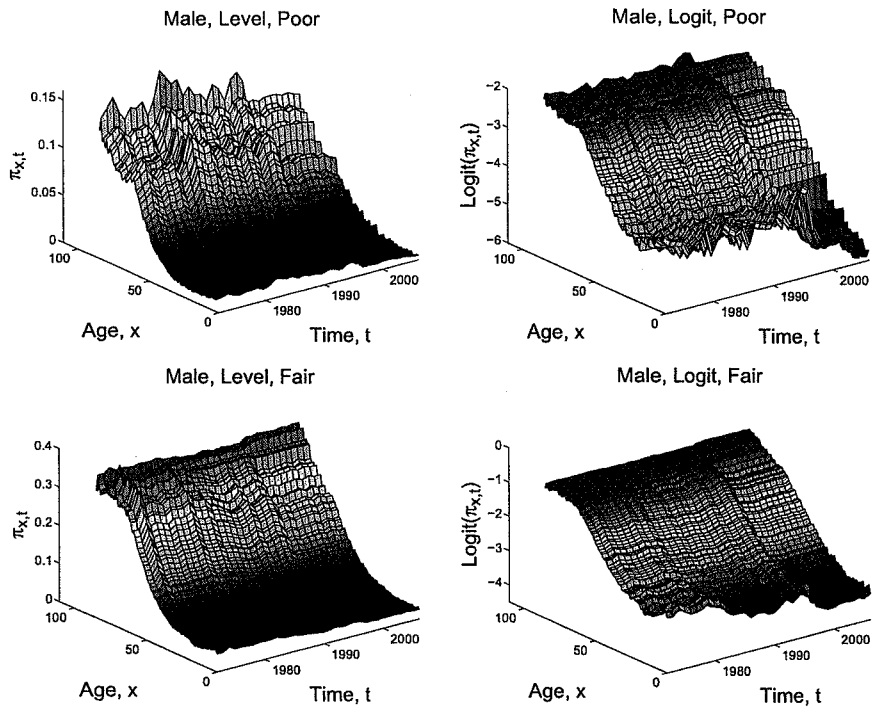

Figure 27: Estimated Health Status Index π_x for Males both in the Level and the Logit Formats

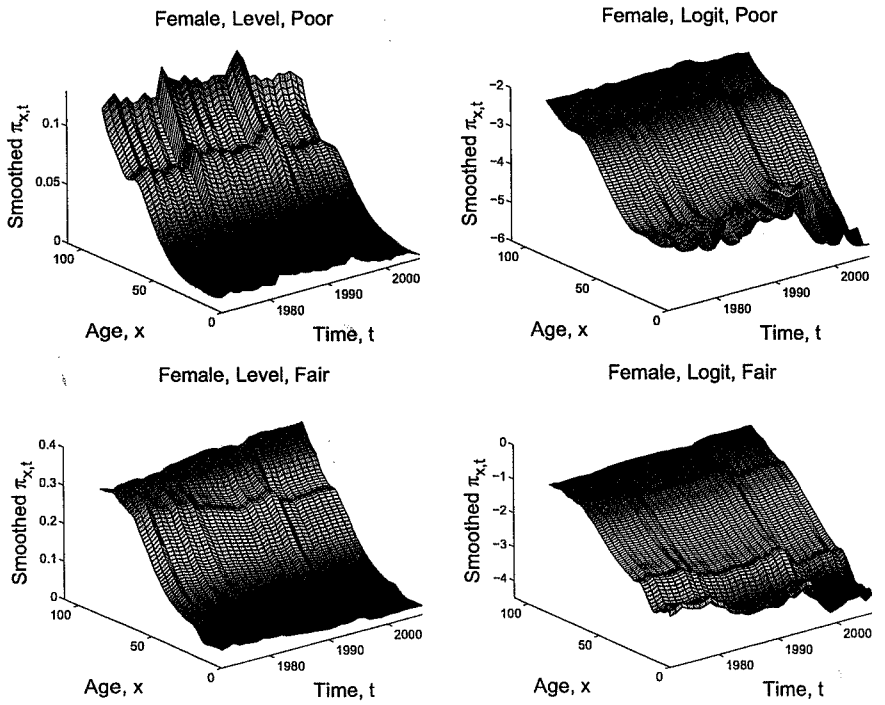

Figure 28: Smoothed Health Status Index π_x for Females both in the Level and the Logit Formats

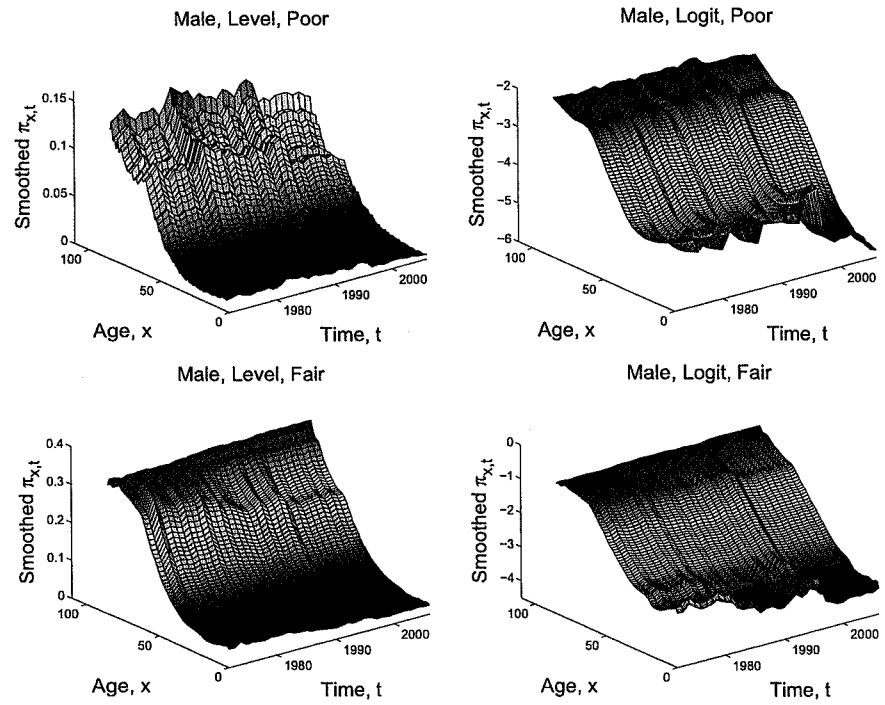

Figure 29: Smoothed Health Status Index π_x for Males both in the Level and the Logit Formats

D.2 Forecasts

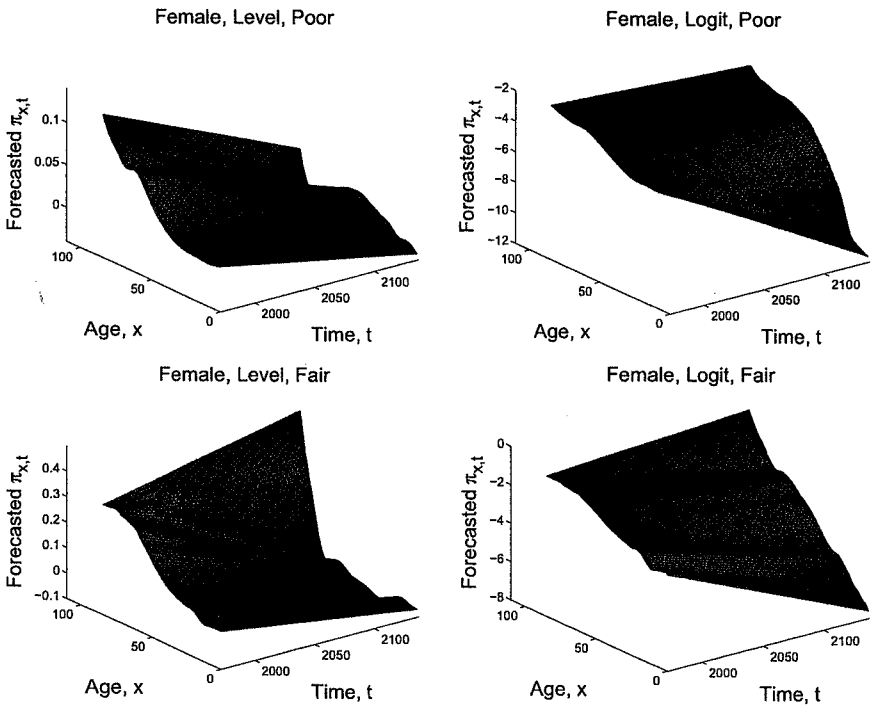

Figure 30: Forecasted Health Status Index π_x for Females both in the Level and the Logit Formats

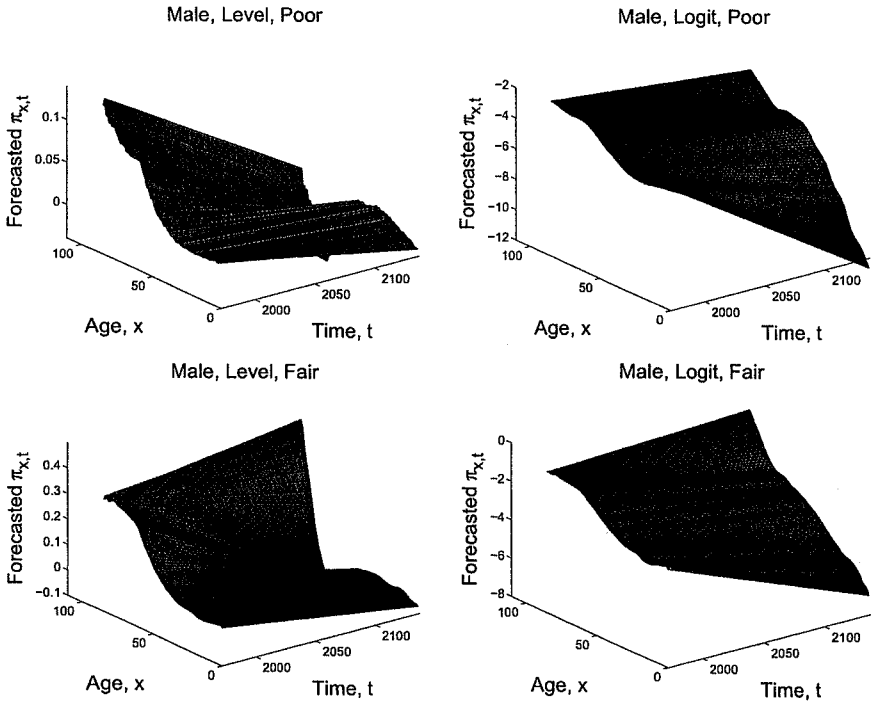

Figure 31: Forecasted Health Status Index π_x for Males both in the Level and the Logit Formats

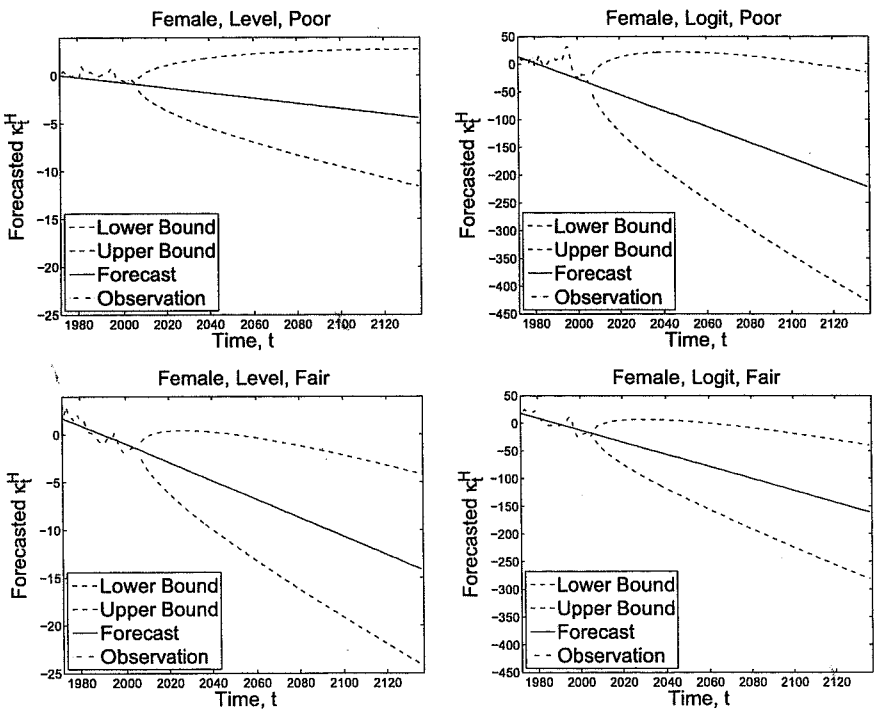

Figure 32: Forecasted κ_x^H for Females and its 95% Confidence Interval

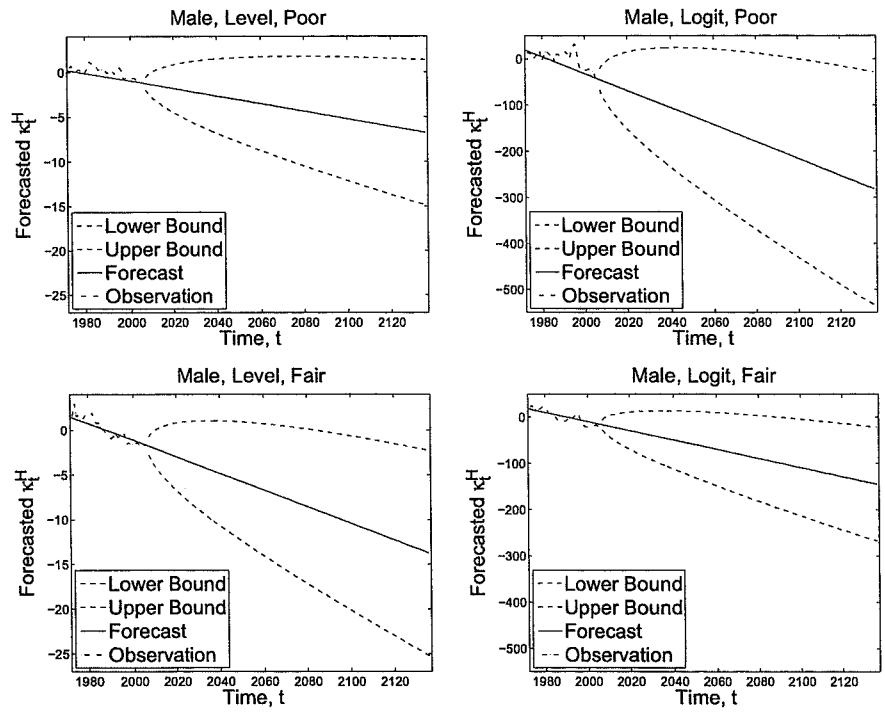

Figure 33: Forecasted κ_x^H for Males and its 95% Confidence Interval

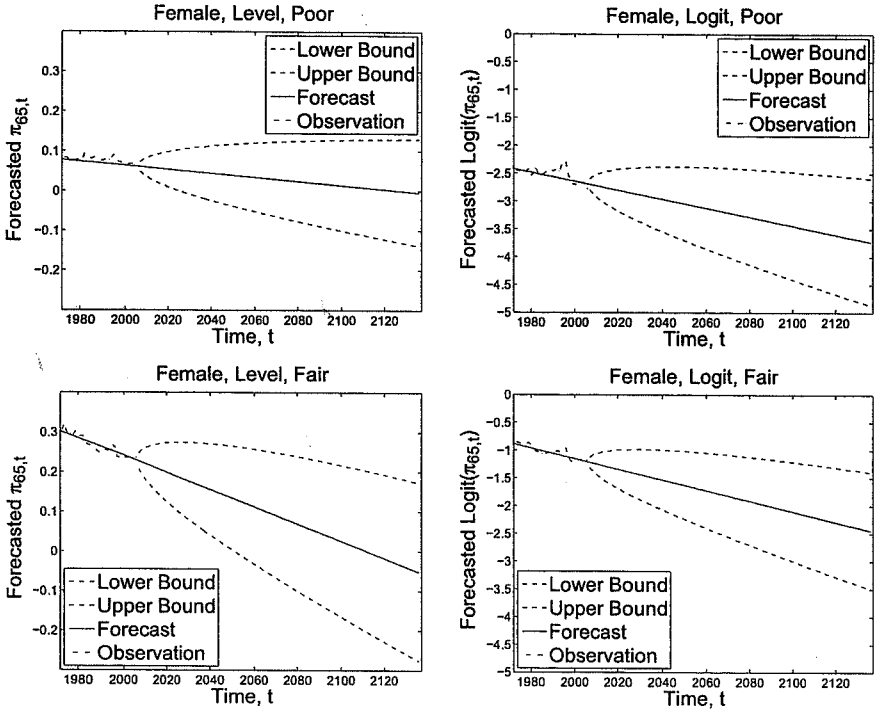

Figure 34: Forecasted (logit)π_x for Females at Age 65 and its 95% Confidence Interval

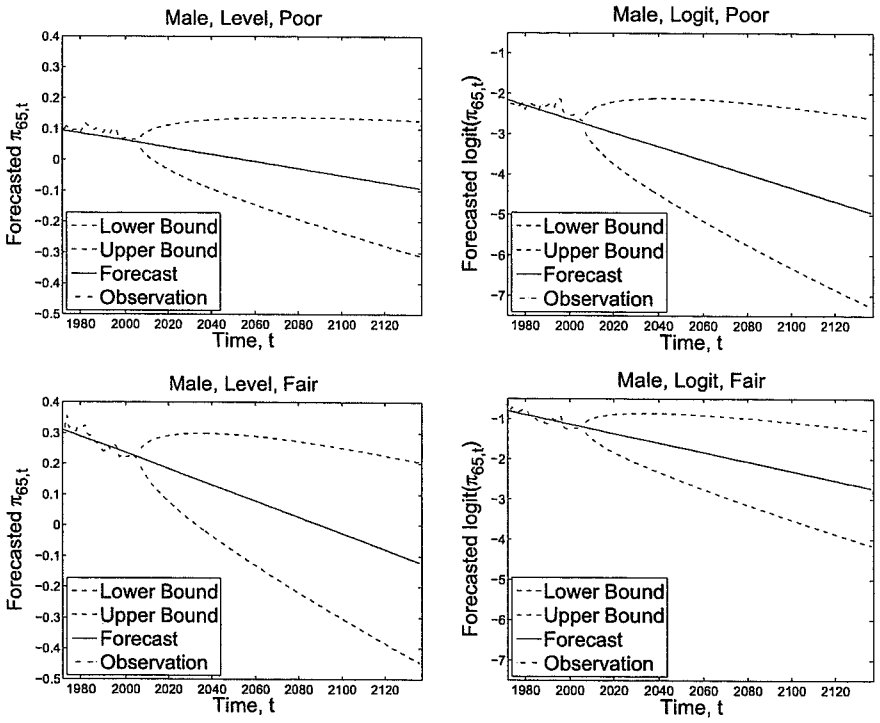

Figure 35: Forecasted (logit)π_x for Males at Age 65 and its 95% Confidence Interval

E Life Expectancy and Healthy Life Expectancy using the Whole Sample

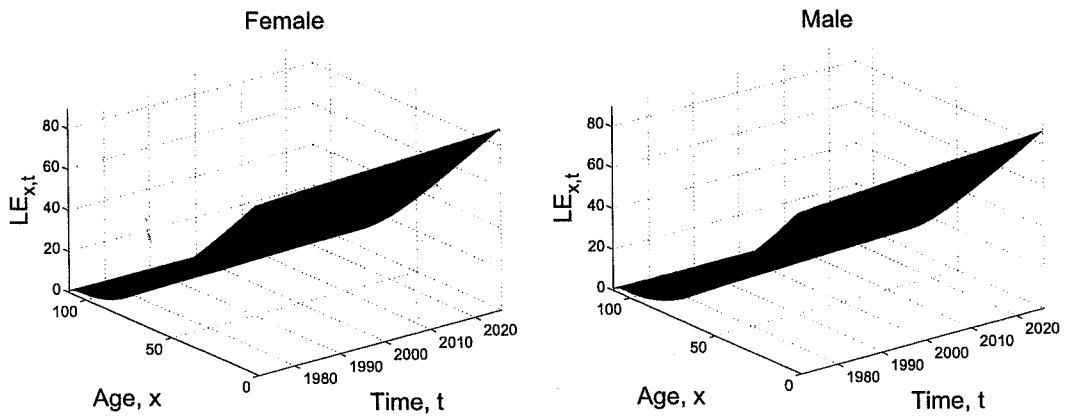

Figure 36: Life Expectancy, $LE_{x,t}$ (Point Estimates and 20 Years Forecasts)

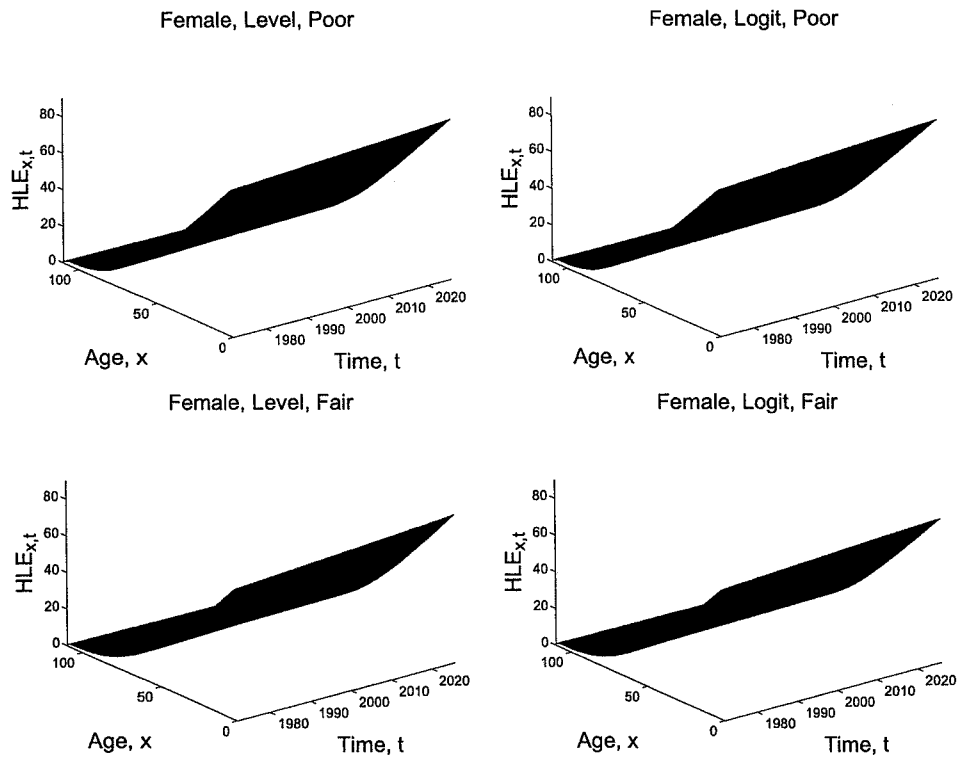

Figure 37: Healthy Life Expectancy ($HLE_{x,t}$) for Females (Point Estimates and Forecasts)

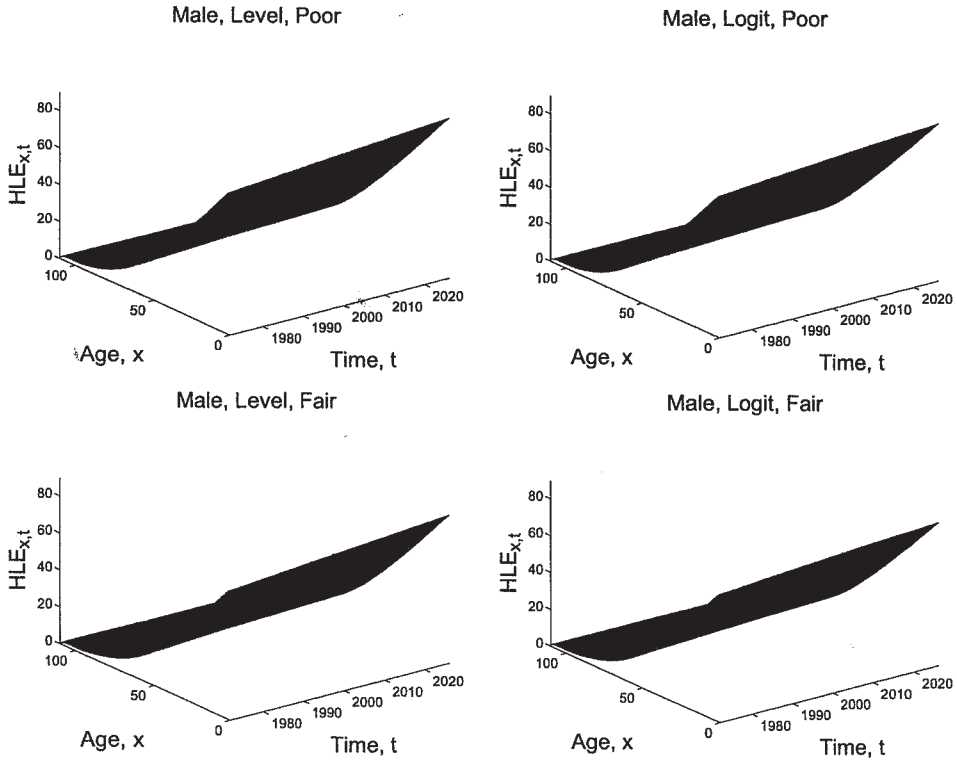

Figure 38: Healthy Life Expectancy ($HLE_{x,t}$) for Males (Point Estimates and Forecasts)

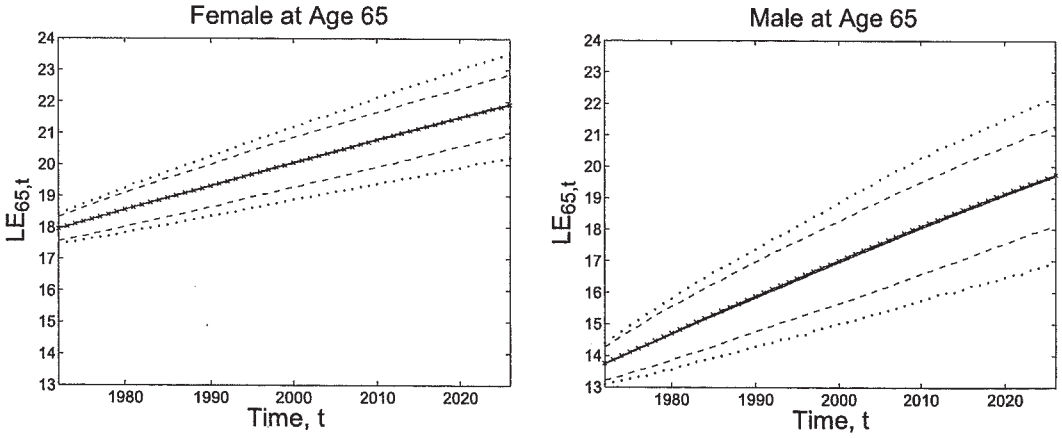

Figure 39: Expected Simulated Life Expectancy at Age 65, $LE_{65,t}$ and its 95% Confidence Interval

'. - . - ' is the point forecasts of LE; '—' denotes the expected simulated LE by only including the process risk of e_t in (31), '- - -' are the corresponding lower and upper bounds; lines '-x-' are the expected simulated LE with process risk (e_t and $\epsilon_{x,t}$) and parameter risk (μ and σ_e),'...' are the corresponding lower and upper bounds.

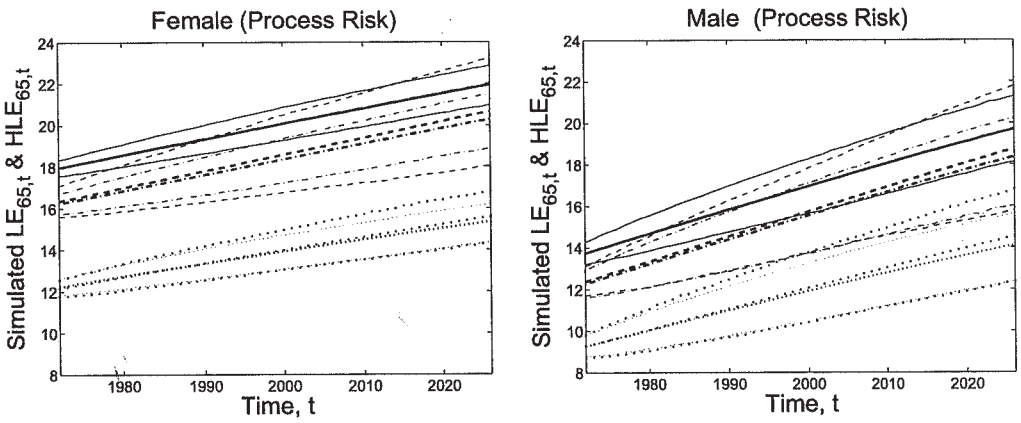

Figure 40: Comparison of Simulated Life Expectancy and Healthy Life Expectancy with Process Risk for Females and Males. '—' denotes LE, '—' are the corresponding 95% confidence interval; '- - -' denotes HLE when HSI cutting ar poor with level format, '- - -' are the corresponding 95% confidence interval ; '. - . -' denotes HLE when HSI cutting ar poor with logit format, '. - . -' are the corresponding 95% confidence interval; '. . .' denotes HLE when HSI cutting at fair with level format, '. . .' are the corresponding 95% confidence interval; '...' denotes HLE when HSI cutting at fair with logit format, '...' are the corresponding 95% confidence interval.

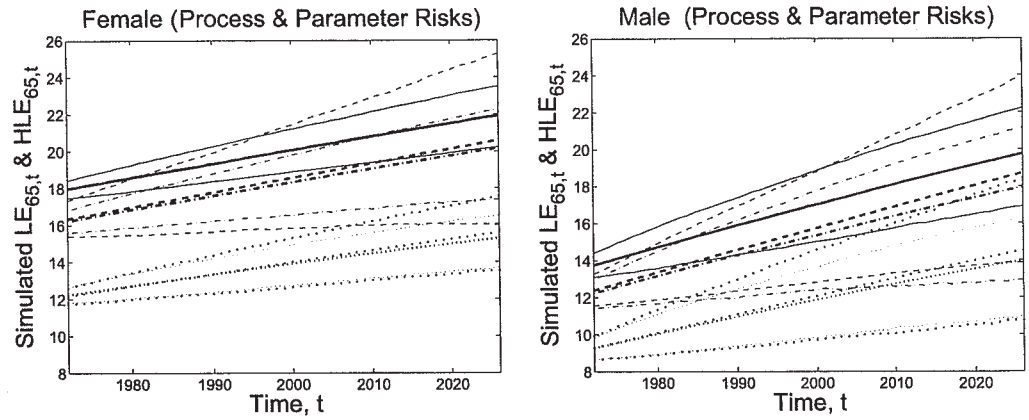

Figure 41: Comparison of Simulated Life Expectancy and Healthy Life Expectancy with both Process and Parameter Risks for Females and Males. '—' denotes LE, '—' are the corresponding 95% confidence interval; '- - -' denotes HLE when HSI cutting ar poor with level format, '- - -' are the corresponding 95% confidence interval ; '. - . -' denotes HLE when HSI cutting ar poor with logit format, '. - . -' are the corresponding 95% confidence interval; '. . .' denotes HLE when HSI cutting at fair with level format, '. . .' are the corresponding 95% confidence interval; '...' denotes HLE when HSI cutting at fair with logit format, '...' are the corresponding 95% confidence interval.

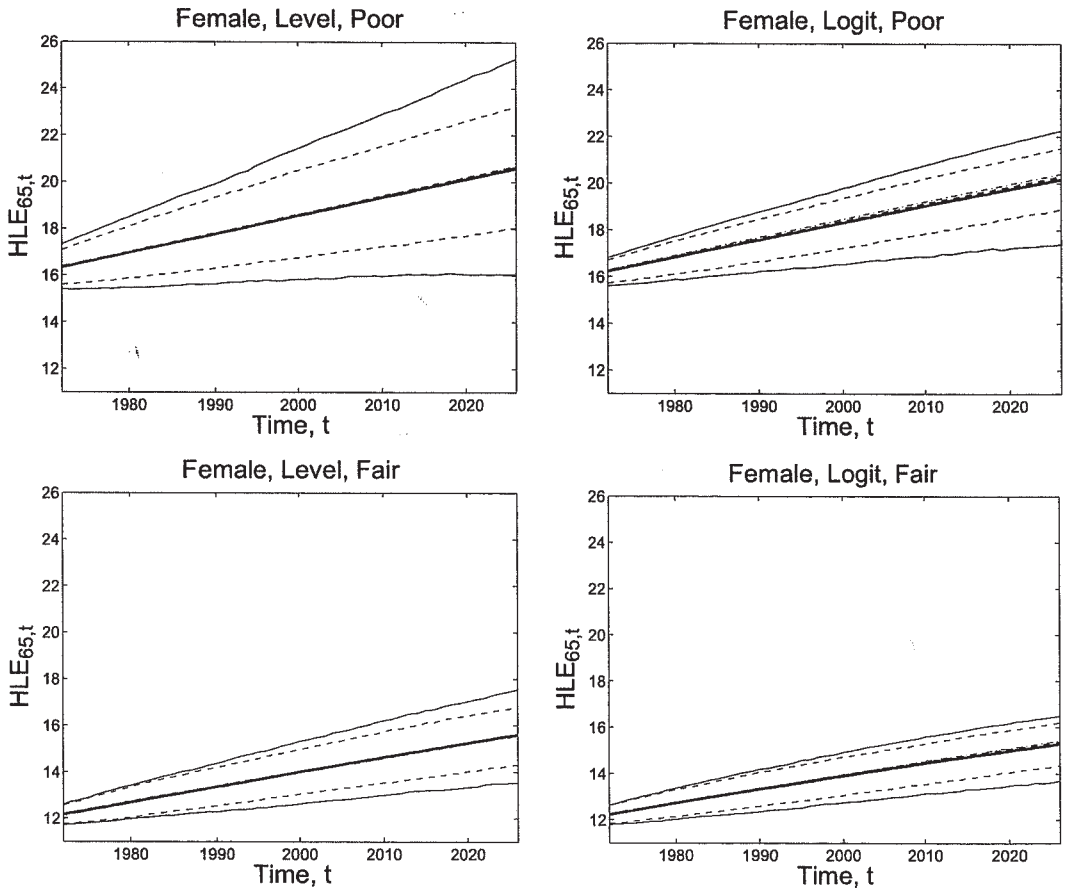

Figure 42: Comparison by Considering Different Risks of Simulated Healthy Life Expectancies for Females at Age 65 with 95% Confidence Intervals. '. - . - ' is the point estimates and forecasts of HLE; '- - -' denotes the expected simulated HLE by only including the process risk of e_t in (31), '- - -' are the corresponding lower and upper bounds; lines '—' are the expected simulated HLE with process risk (e_t and $\epsilon_{x,t}$) and parameter risk (μ and σ_e),'—' are the corresponding lower and upper bounds.

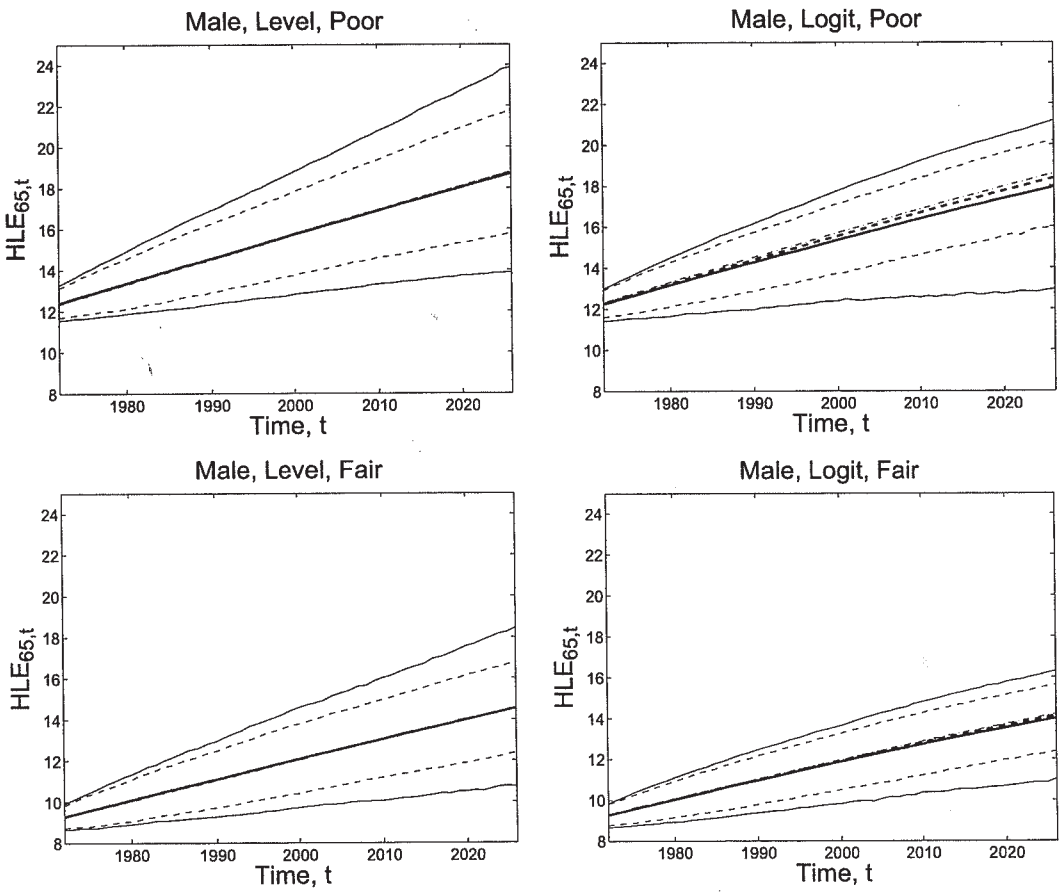

Figure 43: Comparison by Considering Different Risks of Simulated Healthy Life Expectancies for Males at Age 65 with 95% Confidence Intervals. '. - . - ' is the point estimates and forecasts of HLE; '- - -' denotes the expected simulated HLE by only including the process risk of e_t in (31), '- - -' are the corresponding lower and upper bounds; lines '—' are the expected simulated HLE with process risk (e_t and $\epsilon_{x,t}$) and parameter risk (μ and σ_e),'—' are the corresponding lower and upper bounds.

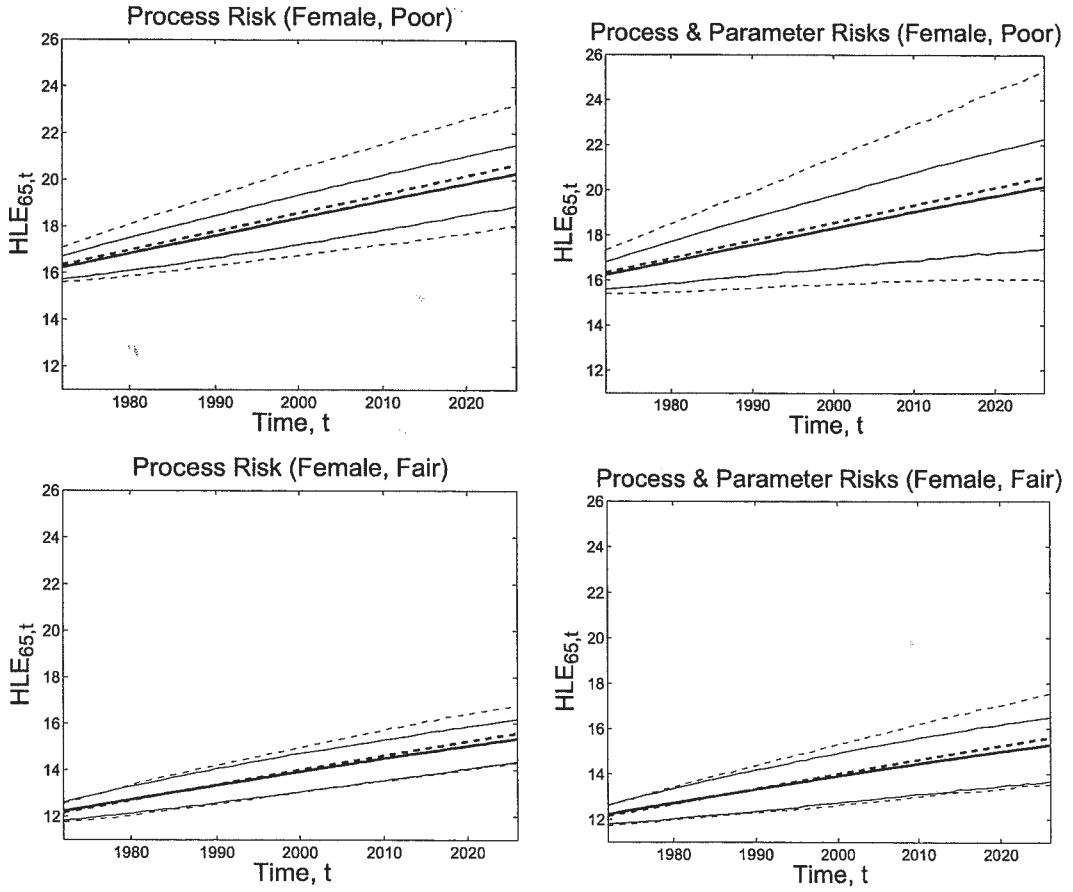

Figure 44: Comparison by Different Health Status Index Formats of Simulated Healthy Life Expectancy for Females at Age 65 with 95% Confidence Interval. '- - -' denotes the expected simulated HLE using level data, '- - -' are the corresponding lower and upper bounds; '—' denotes the expected simulated HLE using logit data, '—' are the corresponding lower and upper bounds.

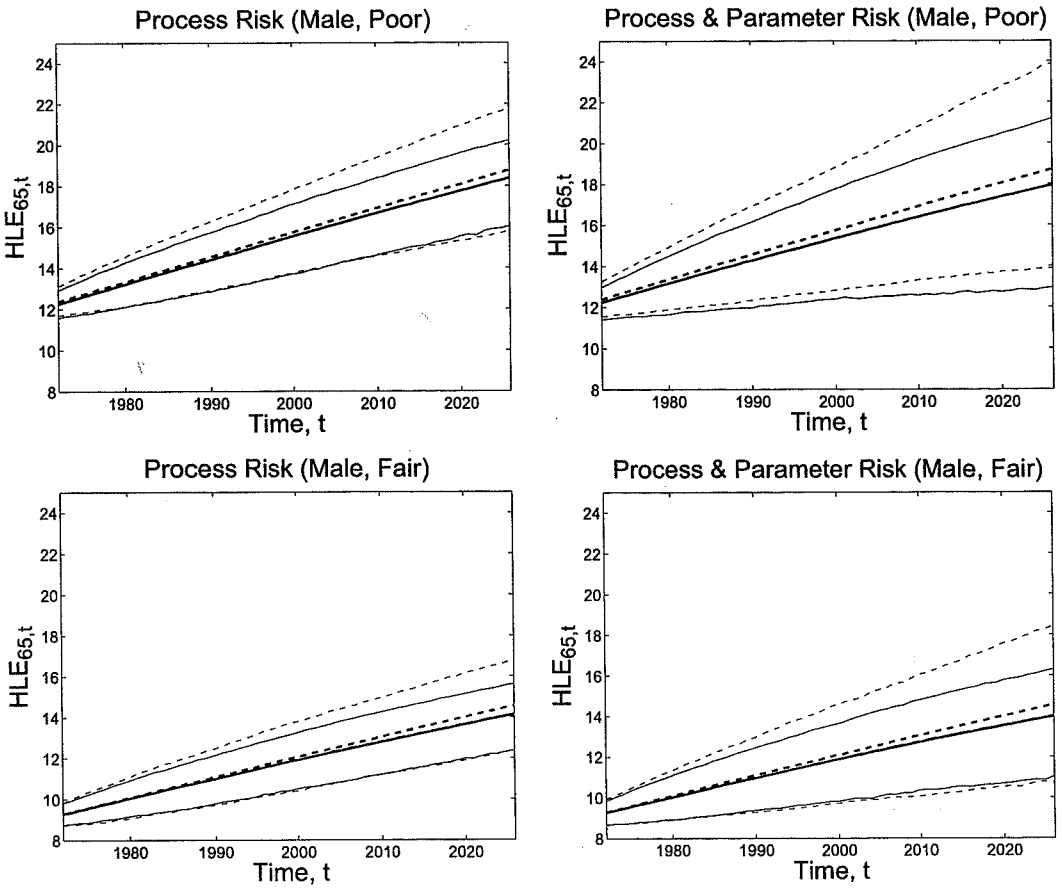

Figure 45: Comparison by Different Health Status Index Formats of Simulated Healthy Life Expectancy for Males at Age 65 with 95% Confidence Interval. '- - -' denotes the expected simulated HLE using level data, '- - -' are the corresponding lower and upper bounds; '—' denotes the expected simulated HLE using logit data, '—' are the corresponding lower and upper bounds.

F Mortality Estimates and Forecasts using the Subsample

F.1 Estimates

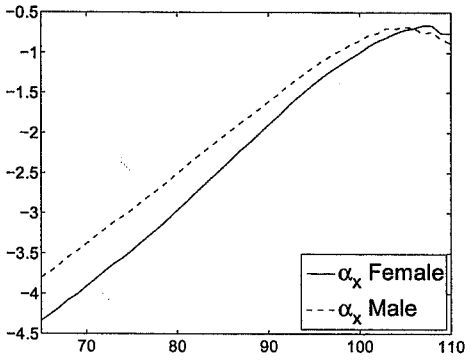

Figure 46: Estimated $\hat{\alpha}_x$ for Mortality Rate

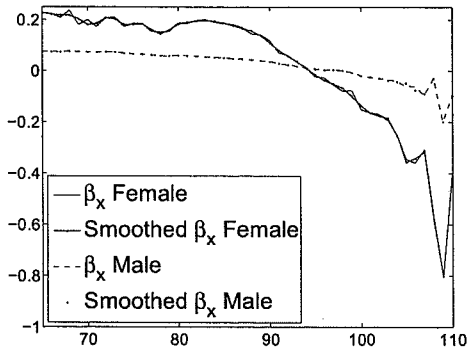

Figure 47: Estimated and Smoothed $\hat{\beta}_x$ for Mortality Rate

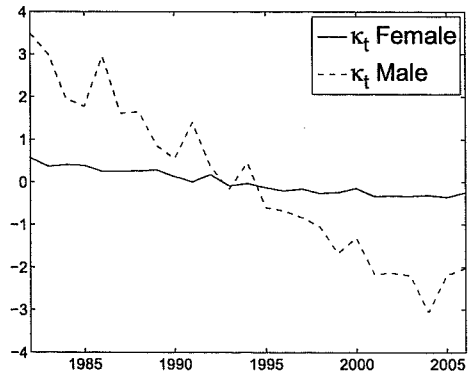

Figure 48: Estimated $\hat{\kappa}_x$ for Mortality Rate

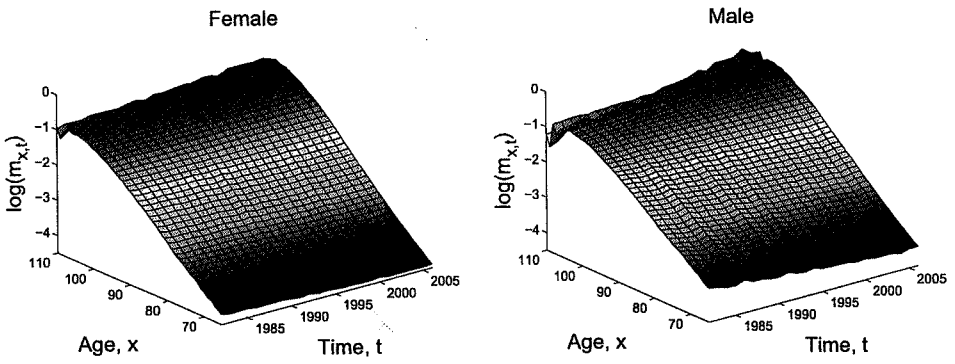

Figure 49: Estimated Log Mortality Rate, $log(m_{x,t})$

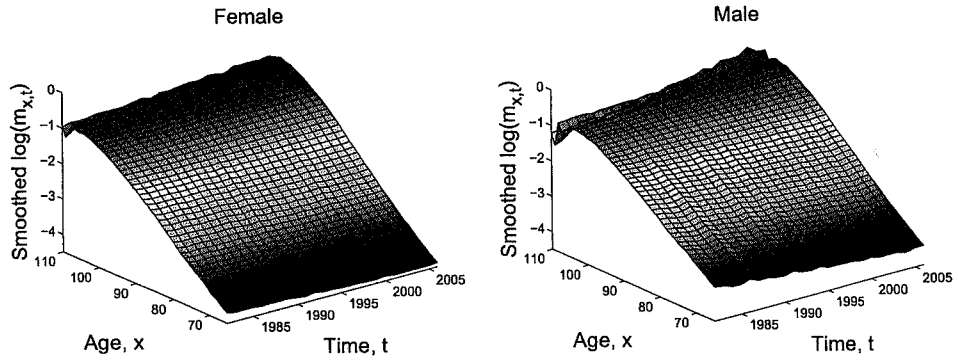

Figure 50: Smoothed Log Mortality Rate, $log(m_{x,t})$

F.2 Forecasts

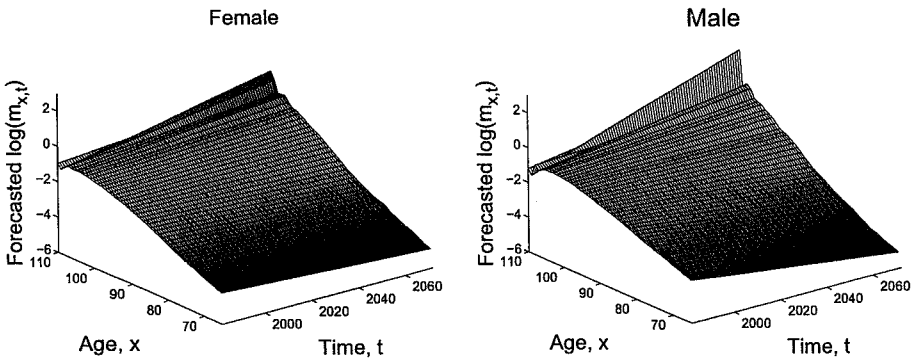

Figure 51: Forecasted Log Mortality Rate, $log(m_{x,t})$

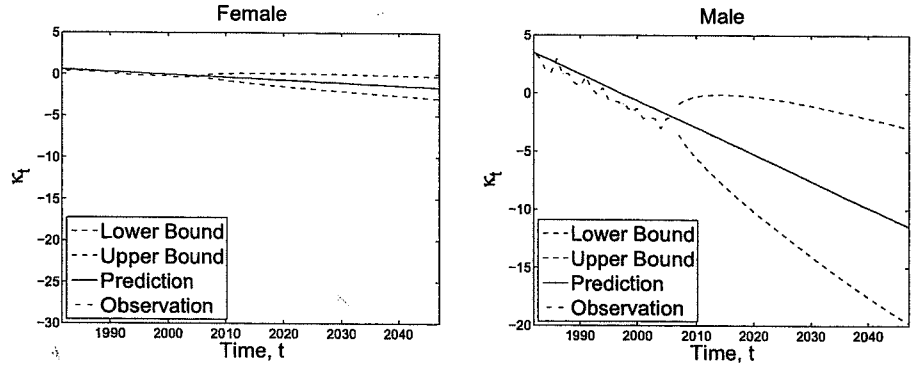

Figure 52: Forecasted κ_x and its 95% Confidence Interval

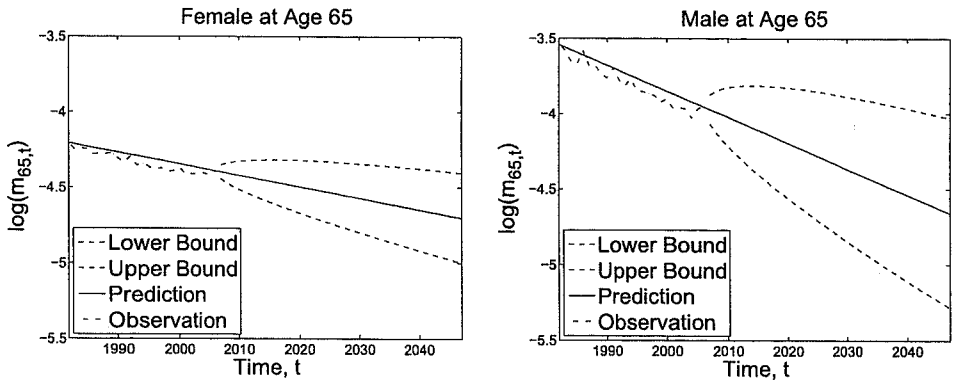

Figure 53: Forecasted Log Mortality Rate at Age 65 ($log(m_{65,t})$) and its 95% Confidence Interval

G Health Estimates and Forecasts using the Subsample

G.1 Estimates

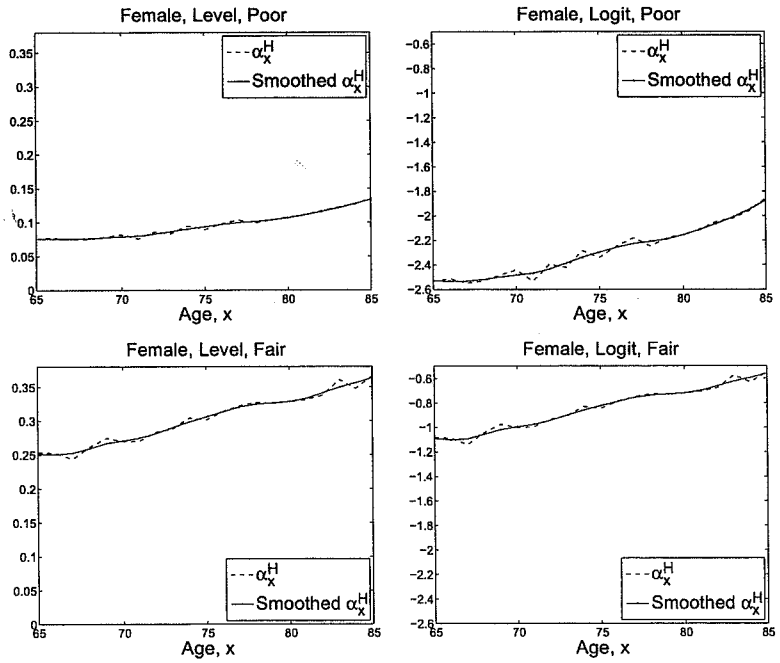

Figure 54: Estimated and Smoothed α_x^H for Female Health Status Index

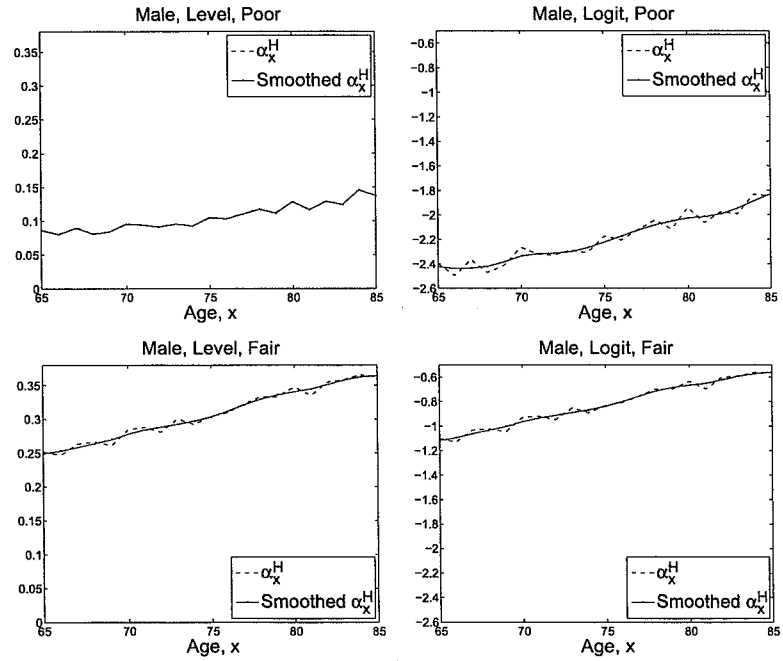

Figure 55: Estimated and Smoothed α_x^H for Male Health Status Index

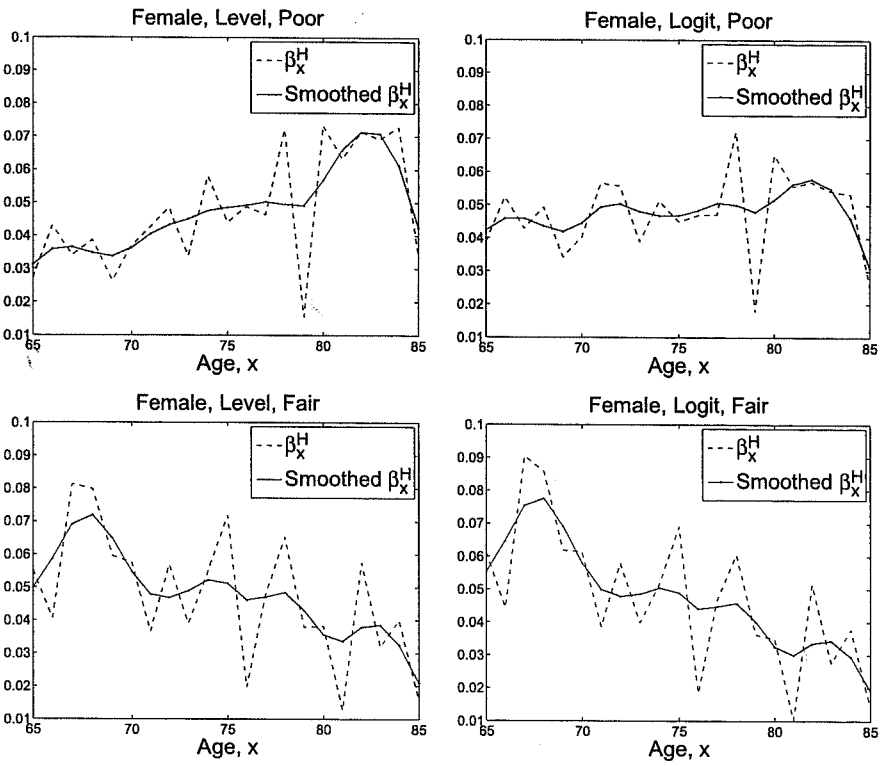

Figure 56: Estimated and Smoothed β_x^H for Female Health Status Index

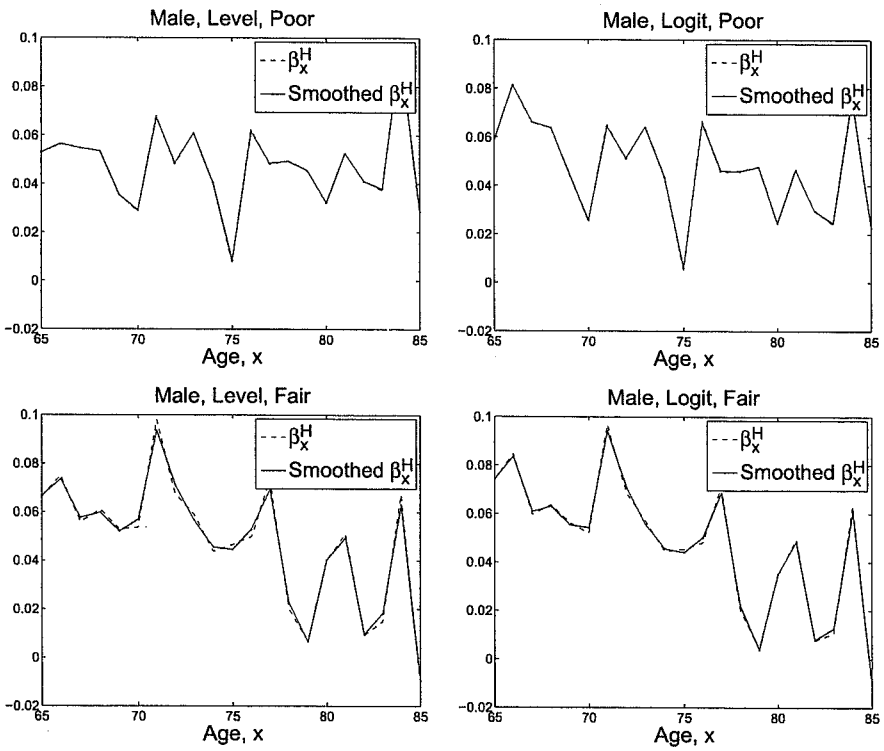

Figure 57: Estimated and Smoothed β_x^H for Male Health Status Index

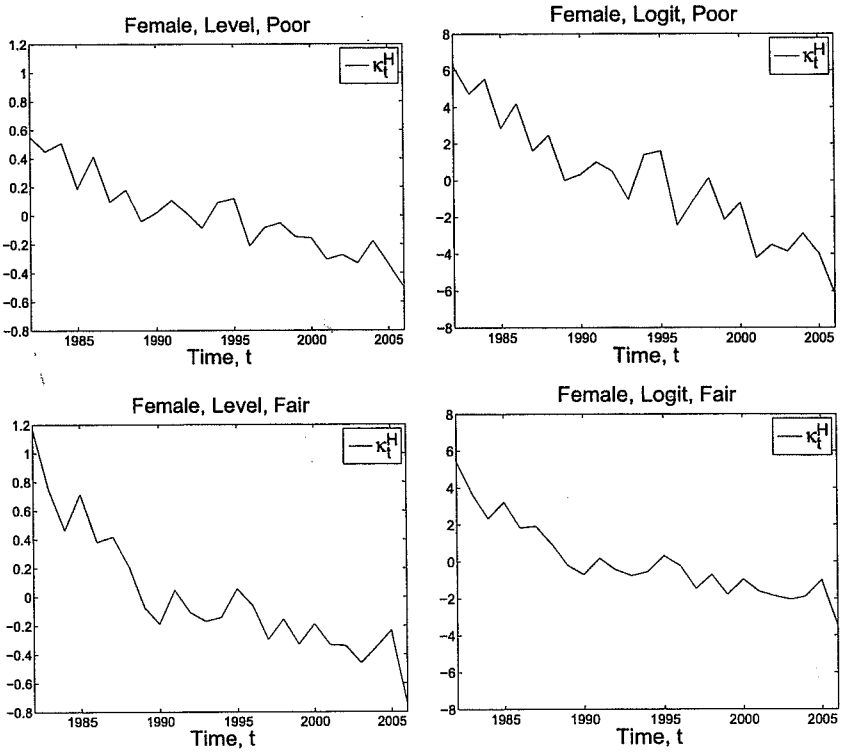

Figure 58: Estimated κ_x^H for Female Health Status Index

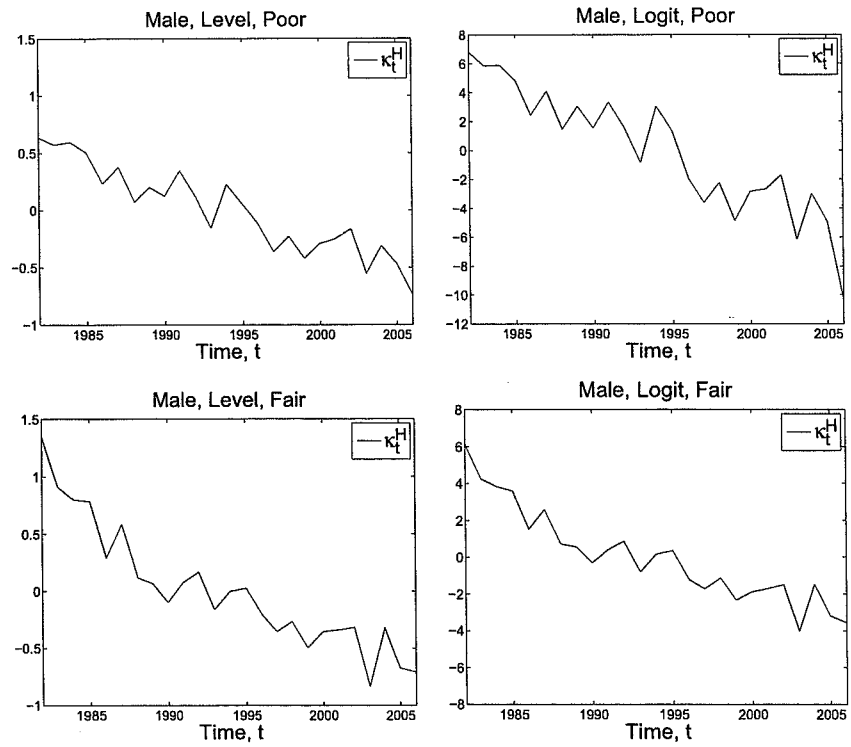

Figure 59: Estimated κ_x^H for Male Health Status Index

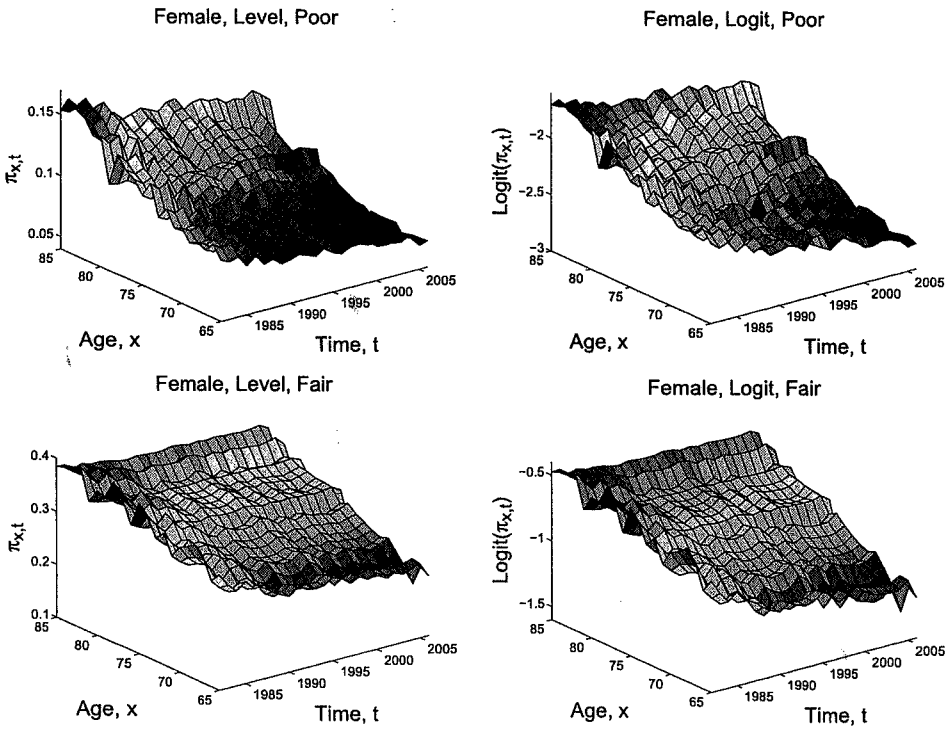

Figure 60: Estimated Health Status Index π_x for Females both in the Level and the Logit Formats

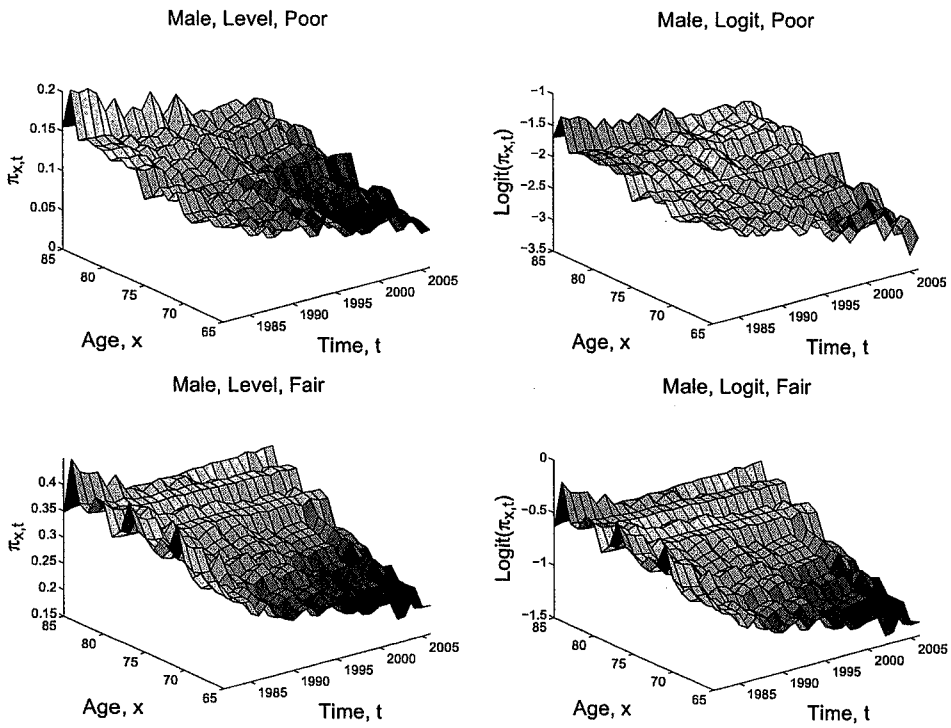

Figure 61: Estimated Health Status Index π_x for Males both in the Level and the Logit Formats

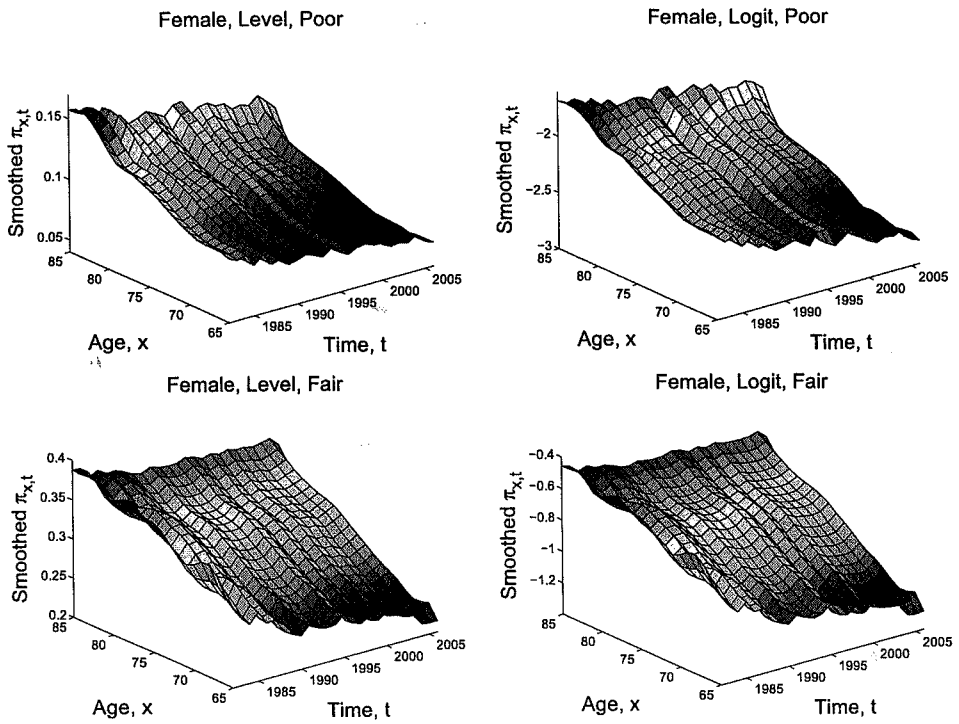

Figure 62: Smoothed Health Status Index π_x for Females both in the Level and the Logit Formats

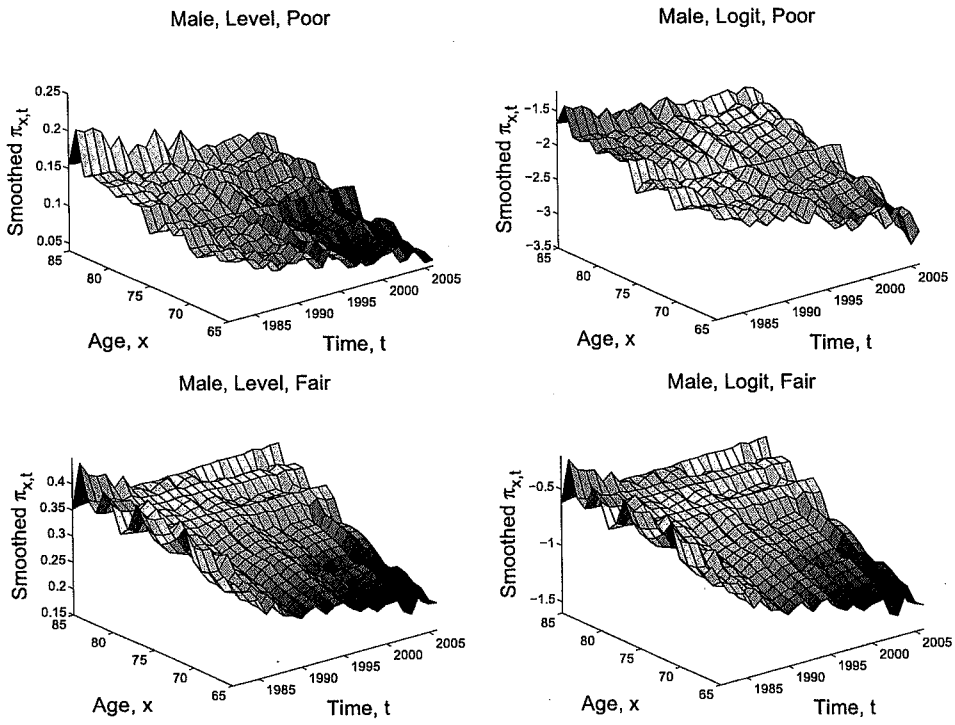

Figure 63: Smoothed Health Status Index π_x for Males both in the Level and the Logit Formats

G.2 Forecasts

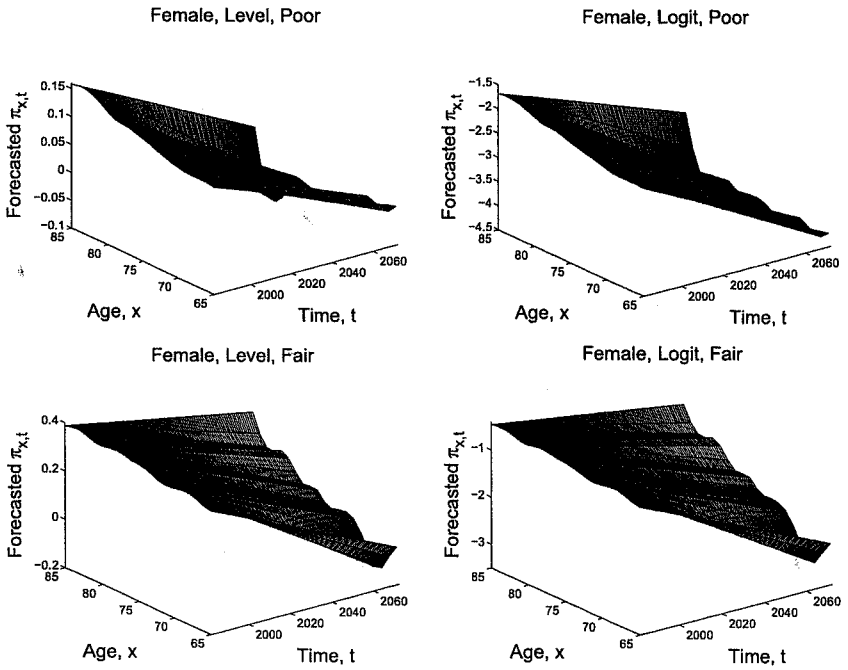

Figure 64: Forecasted Health Status Index π_x for Females both in the Level and the Logit Formats

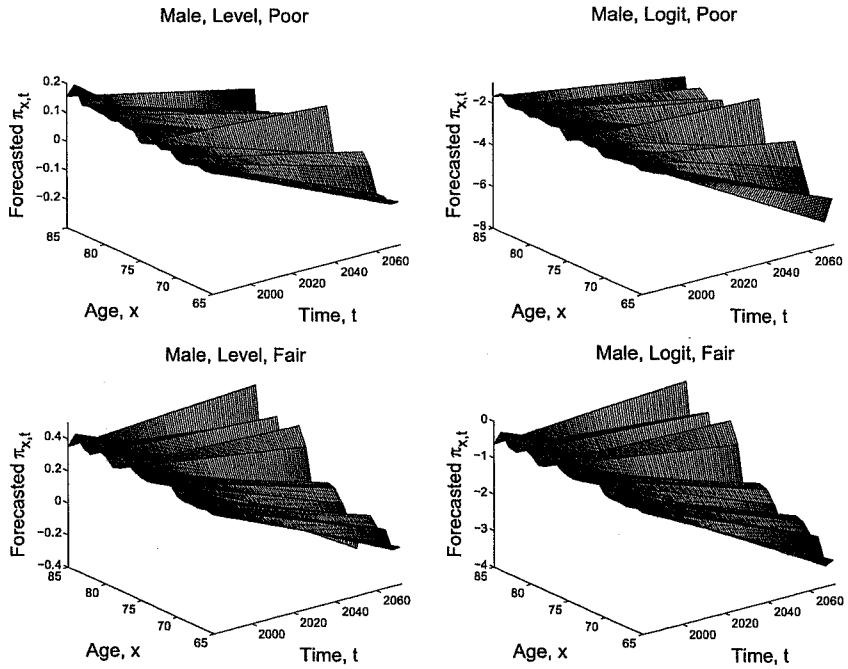

Figure 65: Forecasted Health Status Index π_x for Males both in the Level and the Logit Formats

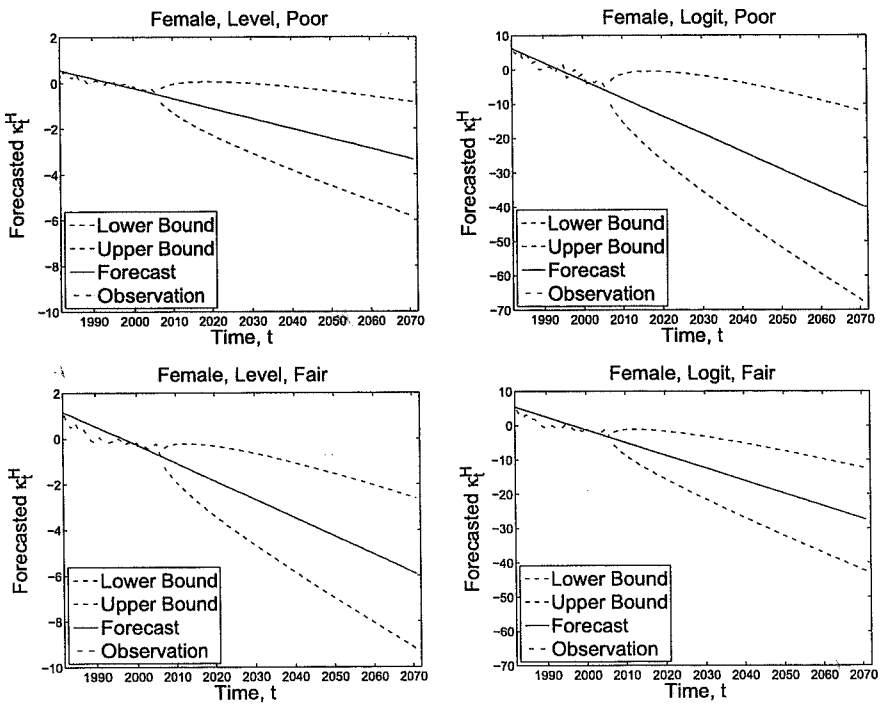

Figure 66: Forecasted κ_x^H for Females and its 95% Confidence Interval

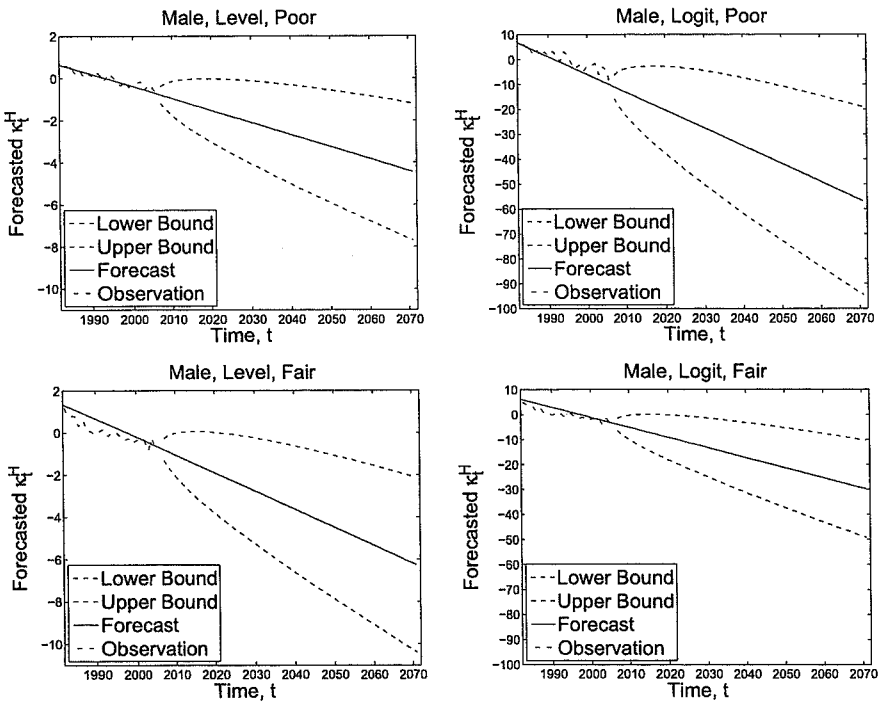

Figure 67: Forecasted κ_x^H for Males and its 95% Confidence Interval

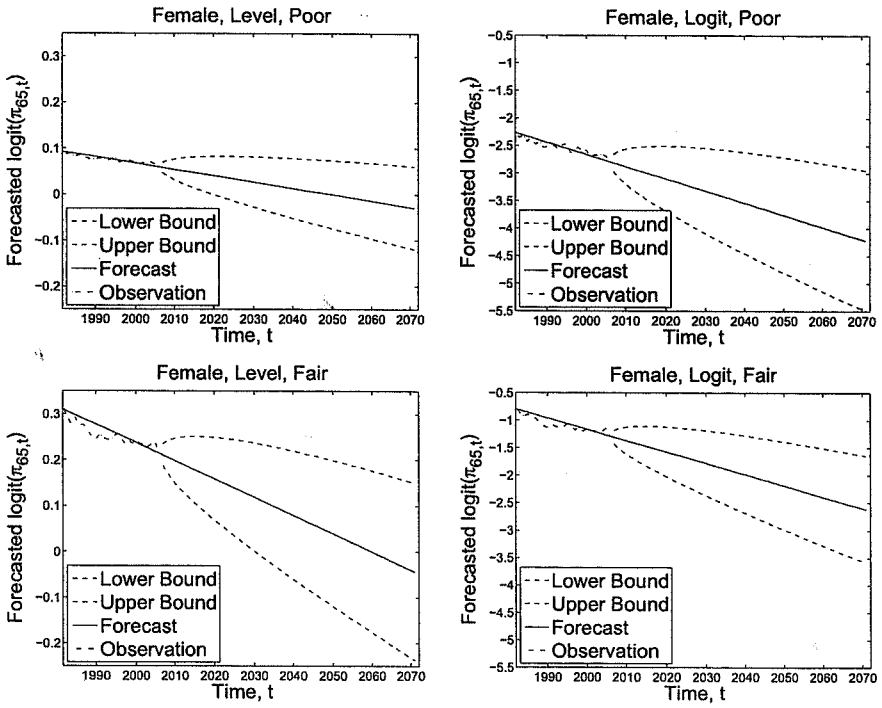

Figure 68: Forecasted (logit)π_x for Females at Age 65 and its 95% Confidence Interval

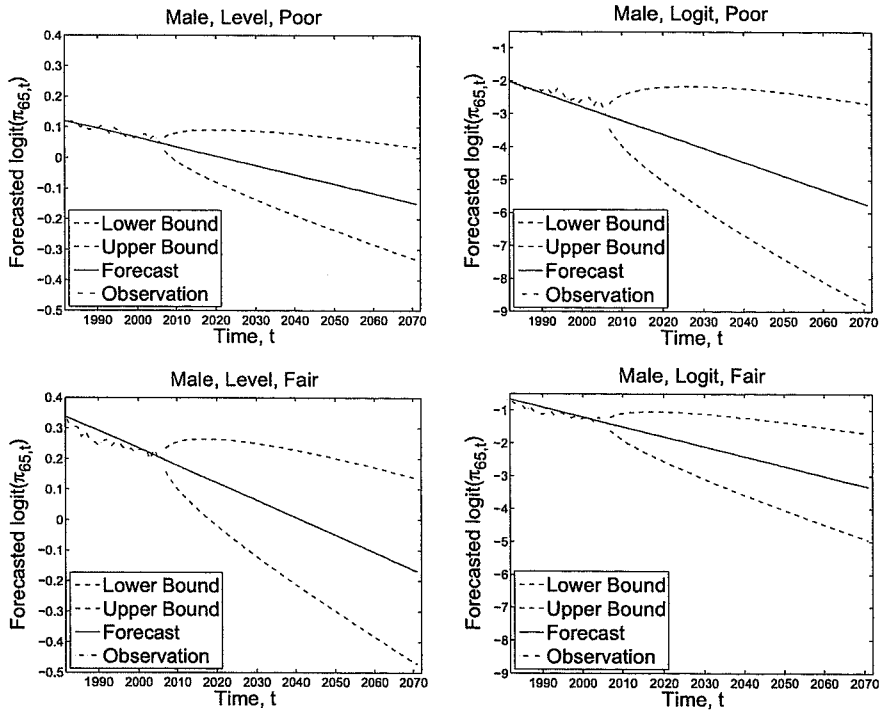

Figure 69: Forecasted (logit)π_x for Males at Age 65 and its 95% Confidence Interval

H Life Expectancy and Healthy Life Expectancy using the Subsample

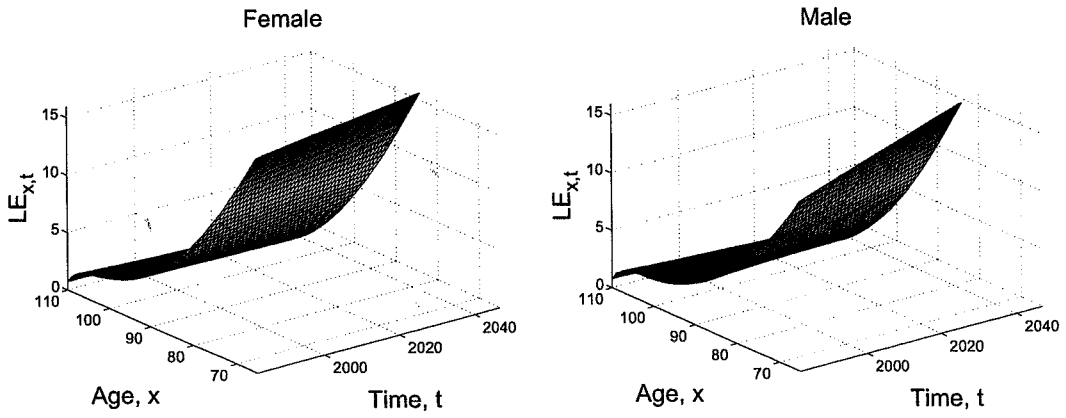

Figure 70: Life Expectancy, $LE_{x,t}$ (Point Estimates and 20 Years Forecasts)

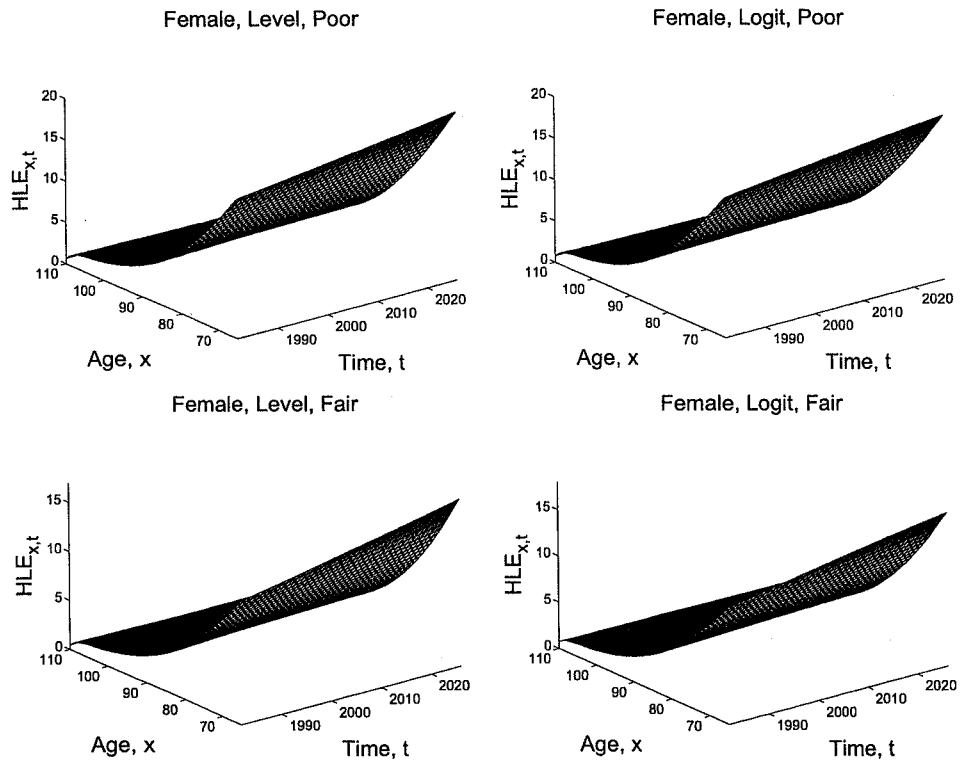

Figure 71: Healthy Life Expectancy ($HLE_{x,t}$) for Females (Point Estimates and Forecasts)

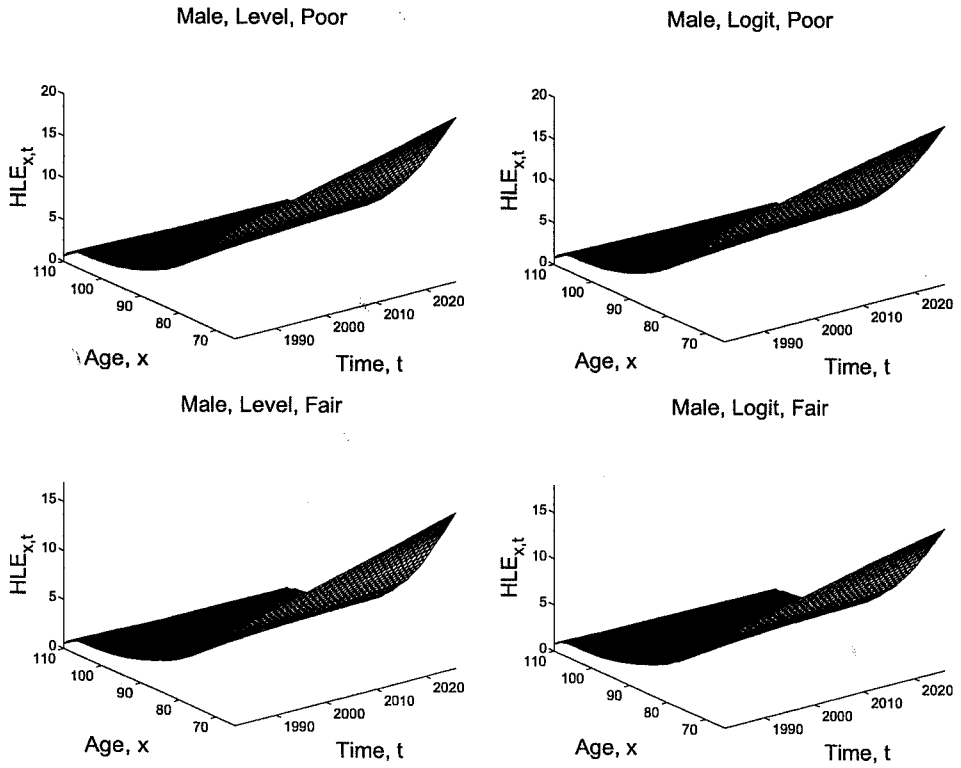

Figure 72: Healthy Life Expectancy ($HLE_{x,t}$) for Males (Point Estimates and Forecasts)

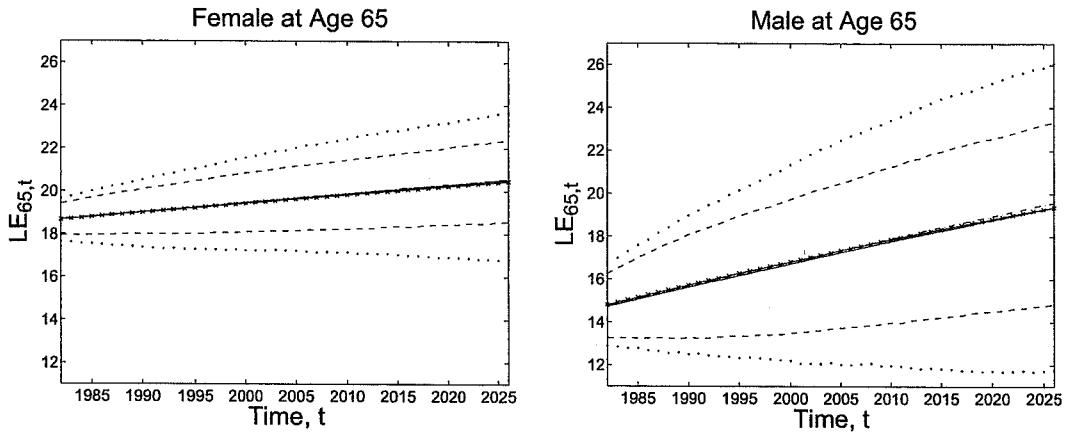

Figure 73: Expected Simulated Life Expectancy at Age 65, $LE_{65,t}$ and its 95% Confidence Interval)
'. - . - ' is the point forecasts of LE; '—' denotes the expected simulated LE by only including the process risk of e_t in (31), '- - -' are the corresponding lower and upper bounds; lines '-x-' are the expected simulated LE with process risk (e_t and $\epsilon_{x,t}$) and parameter risk (μ and σ_e)
,'...' are the corresponding lower and upper bounds.

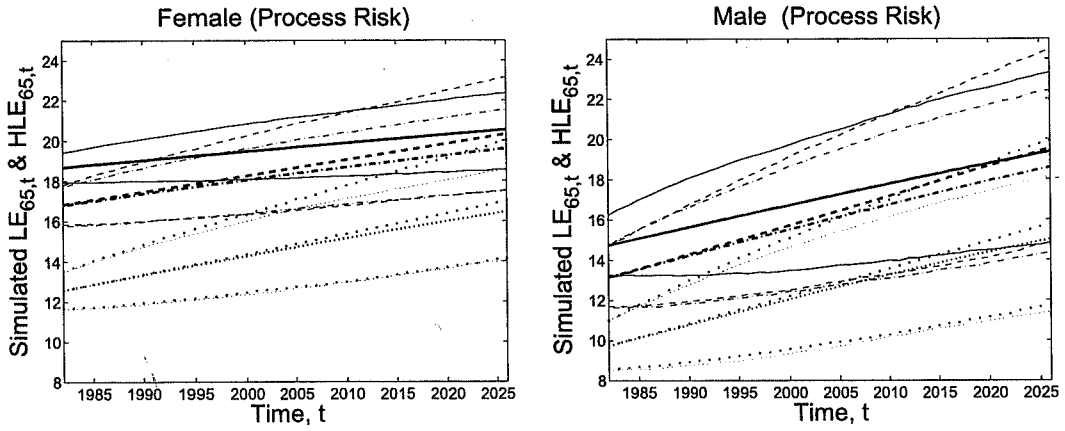

Figure 74: Comparison of Simulated Life Expectancy and Healthy Life Expectancy with Process Risk for Females and Males. '—' denotes LE, '—' are the corresponding 95% confidence interval; '- - -' denotes HLE when HSI cutting ar poor with level format, '- - -' are the corresponding 95% confidence interval ; '. - . -' denotes HLE when HSI cutting ar poor with logit format, '. - . -' are the corresponding 95% confidence interval; '. . .' denotes HLE when HSI cutting at fair with level format, '. . .' are the corresponding 95% confidence interval; '...' denotes HLE when HSI cutting at fair with logit format, '...' are the corresponding 95% confidence interval.

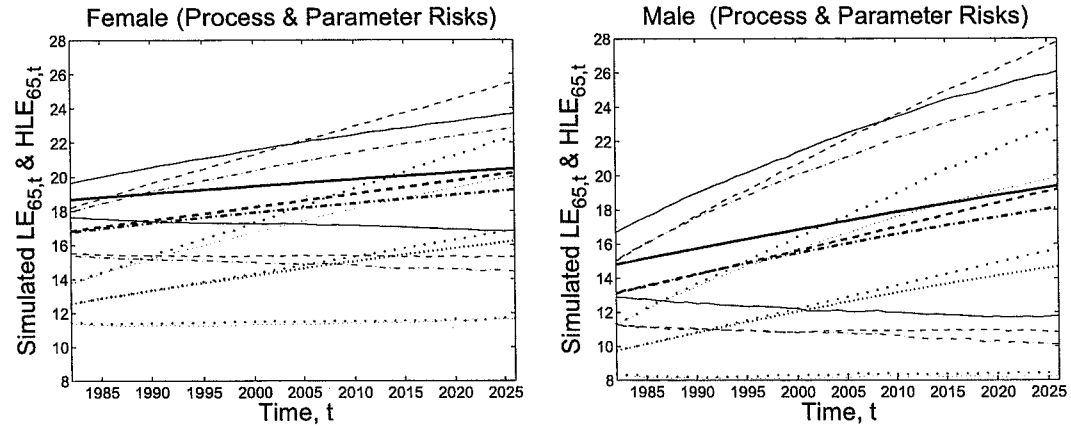

Figure 75: Comparison of Simulated Life Expectancy and Healthy Life Expectancy with both Process and Parameter Risks for Females and Males. '—' denotes LE, '—' are the corresponding 95% confidence interval; '- - -' denotes HLE when HSI cutting ar poor with level format, '- - -' are the corresponding 95% confidence interval ; '. - . -' denotes HLE when HSI cutting ar poor with logit format, '. - . -' are the corresponding 95% confidence interval; '. . .' denotes HLE when HSI cutting at fair with level format, '. . .' are the corresponding 95% confidence interval; '...' denotes HLE when HSI cutting at fair with logit format, '...' are the corresponding 95% confidence interval.

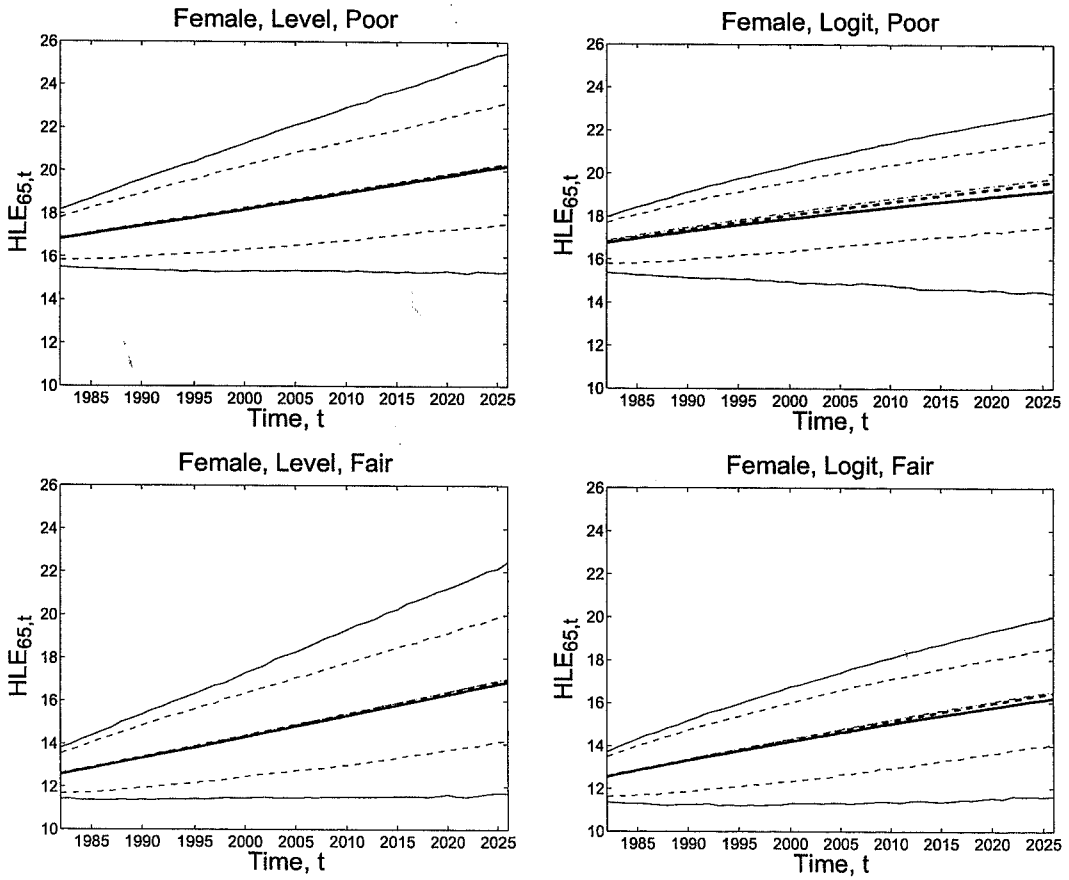

Figure 76: Comparison by Considering Different Risks of Simulated Healthy Life Expectancies for Females at Age 65 with 95% Confidence Intervals. '. - . - ' is the point estimates and forecasts of HLE; '- - -' denotes the expected simulated HLE by only including the process risk of e_t in (31), '- - -' are the corresponding lower and upper bounds; lines '—' are the expected simulated HLE with process risk (e_t and $\epsilon_{x,t}$) and parameter risk (μ and σ_e),'—' are the corresponding lower and upper bounds.

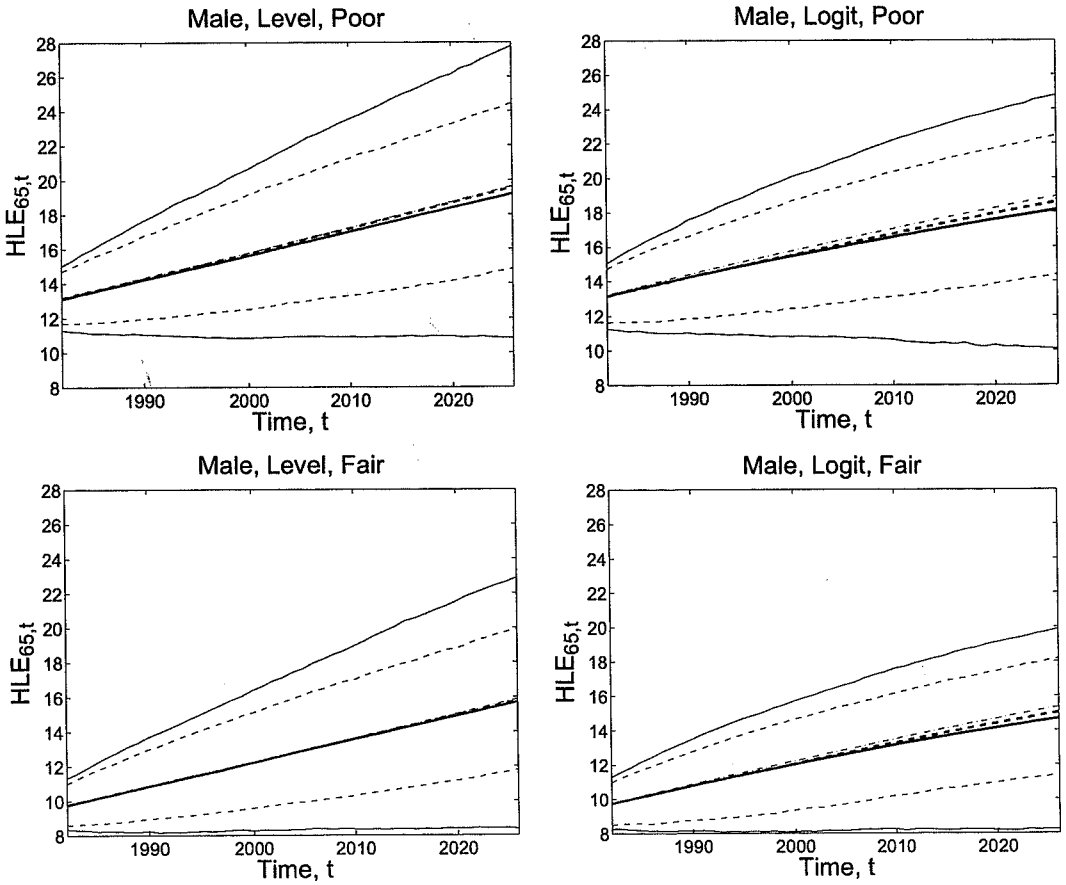

Figure 77: Comparison by Considering Different Risks of Simulated Healthy Life Expectancies for Males at Age 65 with 95% Confidence Intervals. '. - . - ' is the point estimates and forecasts of HLE; '- - -' denotes the expected simulated HLE by only including the process risk of e_t in (31), '- - -' are the corresponding lower and upper bounds; lines '—' are the expected simulated HLE with process risk (e_t and $\epsilon_{x,t}$) and parameter risk (μ and σ_e),'—' are the corresponding lower and upper bounds.

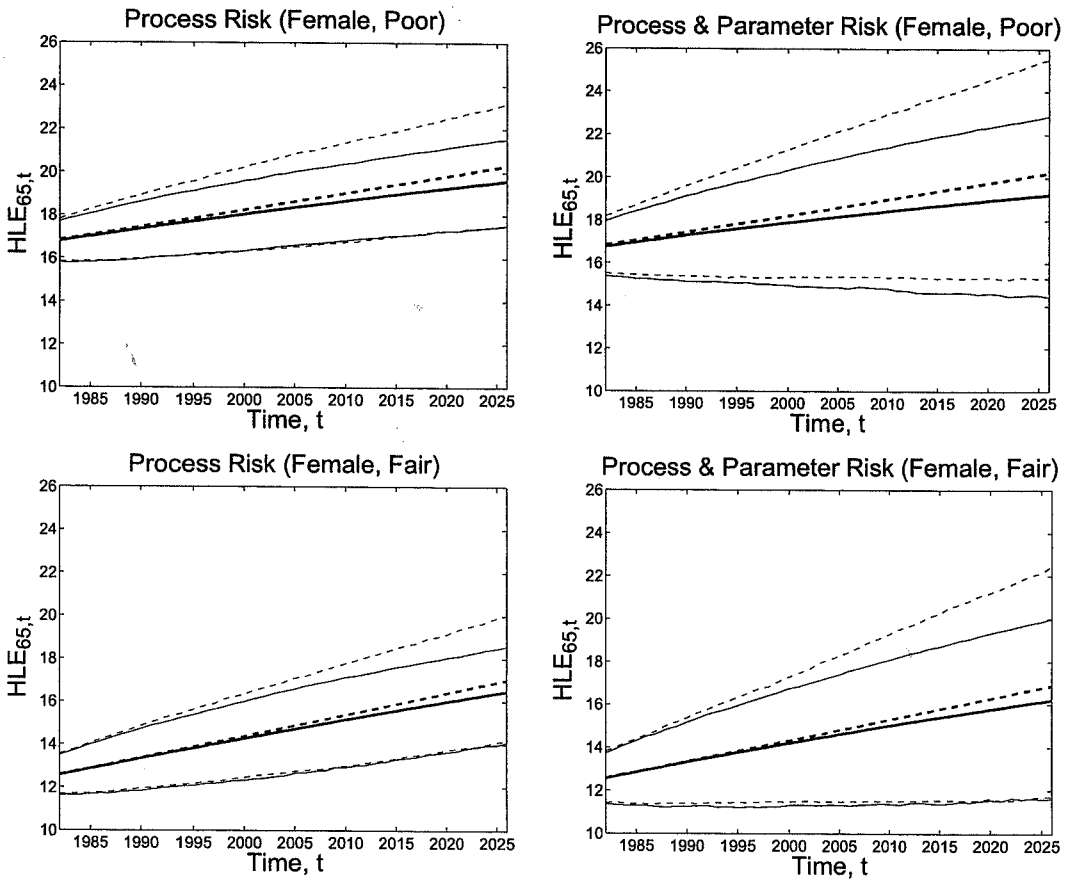

Figure 78: Comparison by Different Health Status Index Formats of Simulated Healthy Life Expectancy for Females at Age 65 with 95% Confidence Interval. '- - -' denotes the expected simulated HLE using level data, '- - -' are the corresponding lower and upper bounds; '—' denotes the expected simulated HLE using logit data, '—' are the corresponding lower and upper bounds.

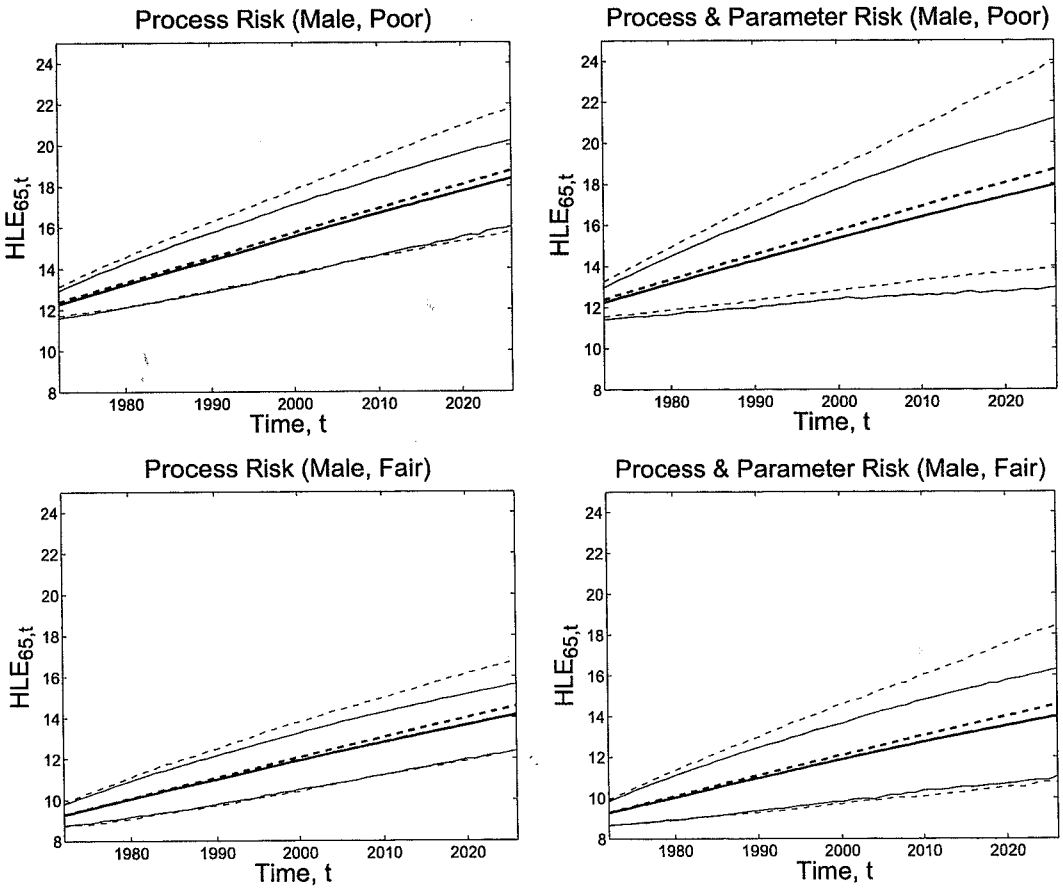

Figure 79: Comparison by Different Health Status Index Formats of Simulated Healthy Life Expectancy for Males at Age 65 with 95% Confidence Interval. '- - -' denotes the expected simulated HLE using level data, '- - -' are the corresponding lower and upper bounds; '—' denotes the expected simulated HLE using logit data, '—' are the corresponding lower and upper bounds.