# ARC Centre of Excellence in Population Ageing Research

# Working Paper 2011/3

## Heterogeneity of Australian Population Mortality and Implications for a Viable Life Annuity Market

Shu Su and Michael Sherris*

* Su is a University of New South Wales (UNSW) Actuarial Studies graduate. Sherris is Professor of Actuarial Studies at UNSW and a Chief Investigator at the ARC Centre of Excellence in Population Ageing Research.

# Heterogeneity of Australian Population Mortality and Implications for a Viable Life Annuity Market

Shu Su
Australian School of Business
University of New South Wales
Email: easonmgid@hotmail.com

Michael Sherris
Australian School of Business
University of New South Wales
Email: m.sherris@unsw.edu.au

7th March 2011

## Abstract

Heterogeneity in mortality rates is known to exist in populations, undermining the use of age and sex as the only rating factors for life insurance and annuity products. Life insurers underwrite life products using a variety of rating factors to allow for this heterogeneity. In the case of life annuities, there is limited underwriting used. Life insurers rely on an assumption that lives will self select and price the longevity risk with an annuity mortality table that assumes above average longevity. This leads to annuities being less attractive to a wide range of individuals, and limits the ability of private annuity markets to meet longevity risk product needs of a large part of the population. There is an increasing use of rating for life annuity pricing such as impaired annuities and postcode underwriting in the UK. In order to fairly price life annuities and support a broader life annuity market, a better understanding of the extent of heterogeneity in population mortality is required. This paper applies well established frailty models and more recently developed Markov models to quantify the extent of heterogeneity in Australian population mortality. The results confirm significant heterogeneity exists. The impact of heterogeneity on life annuity rates and pension costs provides a compelling case for identifying and quantifying more explicitly the factors that determine mortality heterogeneity, particularly at the older ages, including hereditary, socio-economic, and health factors as well as personal habits.

**Keywords:** longevity risk, mortality heterogeneity, frailty model, Markov ageing model, physiological age, annuity pricing

**JEL Classifications:** G22, G23, J11, C46

# 1 Introduction

Heterogeneity of mortality rates is known to exist in populations (Vaupel, Manton, and Stallard (1979) [10]). Although this is taken into account in underwriting life insurance products it is not as common to underwrite life annuity products. Many countries provide social security aged pensions funded through the taxation system and offered on the basis of solidarity with no allowance for risk factors such as age, sex or health status in determining the pension payment. These government aged pensions are usually at a basic level and individuals are either required or encouraged to save for their own retirement through private pensions or other private savings. In the private pensions market the life annuity markets are thin and virtually non-existent in some countries such as Australia (Ganegoda and Bateman (2008) [3]). Adverse selection loadings in premiums, along with capital and risk loadings arising from regulatory requirements such as Solvency II also result in annuity rates that are unattractive to a significant portion of the population.

There are limited studies that quantify the extent to which heterogeneity in population mortality impacts the pricing of life annuities. Olivieri (2006) [9] assesses the risk of a portfolio of life annuities using frailty models but not the implications for life annuity pricing. Individual data required to identify the risk factors contributing to heterogeneity for life annuitants and older aged members of the population is limited because it is confidential information of insurers, confidential individual census data for a population or individual survey data that may not be specifically collected for this purpose. A number of models have been proposed to quantify heterogeneity in population mortality based on widely available population level data. These include frailty models and also a Markov ageing model. A challenge for these models is to separate variability in population mortality rates that arises from heterogeneity as opposed to inherent randomness in mortality.

This paper aims to use Australian population mortality data to assess and fit a number of models of heterogeneity and illustrate the financial impact of heterogeneity by assessing the distribution of life annuity values implied by the models. It is the first study using Australian data to quantify this variability. The paper also compares the different models used and identifies strengths and weaknesses of the models. The results provide a compelling case for identifying more explicitly the factors that determine mortality heterogeneity, particularly at the older ages, including hereditary, socio-economic, and health factors as well as personal habits.

# 2 Frailty Models

Frailty models, introduced in Vaupel et al. (1979), allow for mortality heterogeneity using an unobserved mortality risk factor referred to as frailty, where frailty represents an individuals' relative susceptibility to death compared to a standard. Frailty is assumed to be fixed at birth, and does not vary with age. Frailer individuals are more subject to death, and the survivors on average become less frail as age increases. The selective effect of frailty means that the aging of a cohort as a whole is less than for the standard. With a frailty model, observed mortality at older ages improves relative to the standard because the frailer lives die relatively earlier.

The frailty factor is usually defined in terms of the force of mortality. For an indivi-

dual aged x with frailty z, the force of mortality has the form:

$$\mu(x,z) = z \cdot \mu(x,1)$$

where $\mu(x,1)$ is the standard force of mortality or the force of mortality for individuals with frailty 1. The frailty factor $z$ is unobserved, and is assumed to follow a speciifed statistical distribution.

Under this definition for frailty, there are two sources of variability in observed mortality experience. One comes from the randomness of time to death given the mortality rates $\mu$. The other source comes from the stochastic variability of $\mu$. Individual mortality rates differ because of heterogeneity of the population. Heterogeneity is the source of variability in $\mu$ in frailty models.

The standard force of mortality and the distribution of the frailty factor can not both be directly determined from population mortality data. Assumptions are required for these in order to fit and assess different models. A common assumption is that the standard force of mortality follows a Gompertz mortality function with:

$$\mu(x,1) = \alpha \cdot e^{\beta x} \tag{1}$$

For the frailty distribution, common assumptions include Gamma, Inverse Gaussian, and Lognormal. The Gamma and Inverse Gaussian distributions for frailty will be considered.

The following notation will be used throughout for the frailty model:

$\mu(x,1)$ : standard force of mortality at age x

$\mu(x,z)$ : force of mortality for an individual with frailty z

$\bar{z}_x$ : mean frailty at age x

$f_Z(z)$ : marginal density function of frailty distribution

$f_{Z|X}(z|X=x)$ : density function of frailty for survivors at age x

$f^*_{Z|X}(z|X=x)$ : density function of frailty distribution for deaths at age x

$f_X(x)$ : marginal density function of survival time

$f_{X|Z}(x|Z=z)$ : Conditional density function of survival time given frailty z

$f_{X,Z}(x,z)$ : Joint density function of time to death x and frailty z

$s_{X|Z}(x|Z=z)$ : Conditional survival function given frailty z

$H(x)$ : Cumulative hazard of standard force of mortality at age x

$\mu_x$ : Force of mortality at age x, which is a random variable

$f(\mu_x)$ : Density function of individual force of mortality at age x

$\bar{\mu}_x$ : Mean force of mortality at age x

$\hat{\mu}_x$ : Sample mean of force of mortality at age x, which is a random variable

## 2.1  Gamma Distributed Frailty

Under the assumption of Gamma distributed frailty with shape parameter $k$ and scale parameter $\lambda$ ($Gamma(k,\lambda)$), the marginal density of frailty $f_{Z|X}(z|X=0)$ or $f_Z(z)$ is:

$$f_Z(z) = f_{Z|X}(z|X=0) = \frac{\lambda^k}{\Gamma(k)} \cdot z^{k-1} \cdot e^{-\lambda \cdot z} \tag{2}$$

The mean frailty at birth is $E[z] = \bar{z}_x = k/\lambda$. The level of population heterogeneity is measured by either the variance $\frac{k}{\lambda^2}$ or coefficient of variation $\sqrt{\frac{1}{k}}$. A nice property of assuming a Gamma distributed frailty is that the distribution of frailty at different ages also follows a Gamma distribution with the same shape parameter (Vaupel et al. 1979). That is, conditional on surviving up to age x, the distribution of frailty is $Gamma(k, \lambda(x))$ with density:

$$f_{Z|X}(z|X = x) = \frac{(\lambda(x))^k}{\Gamma(k)} \cdot z^{k-1} \cdot e^{-\lambda(x) \cdot z} \tag{3}$$

where

$$\begin{aligned} \lambda(x) &= \lambda + H(x) \\ &= \lambda + \int_0^x \mu(t) dt \end{aligned} \tag{4}$$

This is shown using the definition of force of mortality in the form of a conditional distribution:

$$\mu(x, z) = \frac{f_{X|Z}(x|Z = z)}{s_{X|Z}(x|Z = z)} \tag{5}$$

where

$$\begin{aligned} s_{X|Z}(x|Z = z) &= e^{-\int_0^x \mu(t,z) dt} \\ &= e^{-z \cdot \int_0^x \mu(t,1) dt} \\ &= e^{-z \cdot H(x)} \end{aligned} \tag{6}$$

Therefore,

$$\begin{aligned} f_{X|Z}(x|Z = z) &= \mu(x, z) \cdot s_{X|Z}(x|Z = z) \\ &= z \cdot \mu(x, 1) \cdot e^{-z \cdot H(x)} \end{aligned} \tag{7}$$

From the relationship between a conditional and an unconditional distribution, the joint distribution of age and frailty $f_{X,Z}(x, z)$ is:

$$\begin{aligned} f_{X,Z}(x, z) &= f_{X|Z}(x|Z = z) \cdot f_Z(z) \\ &= z \cdot \mu(x, 1) \cdot e^{-z \cdot H(x)} \cdot \frac{\lambda^k}{\Gamma(k)} \cdot z^{k-1} \cdot e^{-\lambda \cdot z} \\ &= \frac{\lambda^k}{\Gamma(k)} \cdot z^k \cdot e^{-\lambda(x) \cdot z} \cdot \mu(x, 1) \end{aligned} \tag{8}$$

The conditional distribution of $z$ given survival up to age $x$ is obtained by integrating

the joint density function from 0 to infinity with respect to $x$:

$$f_{Z|X}(z|X = x) = f_{X,Z}(x,z|X > x) = \int_x^\infty f_{T,Z}(t,z)dt$$

$$= \int_x^\infty f_Z(z) \cdot f_{T|Z}(t|Z = z)dt$$

$$= f_Z(z) \cdot \int_x^\infty f_{T|Z}(t|Z = z)dt \tag{9}$$

$$= \frac{\lambda^k}{\Gamma(k)} \cdot z^{k-1} \cdot e^{-\lambda \cdot z} \cdot e^{-z \cdot H(x)}$$

$$= \frac{\lambda^k}{\Gamma(k)} \cdot z^{k-1} \cdot e^{-(\lambda + H(x)) \cdot z} \tag{10}$$

Normalizing the result to make it a density function (integrate to 1) gives the form in (3), which is a $Gamma(k, \lambda(x))$.

The mean frailty of the cohort, $\frac{k}{\lambda(x)}$, is decreasing as age increases. This is the selection effect that is a feature of the frailty model. The lower the value of $k$, the higher the level of heterogeneity, and the faster the decrease in mean frailty of the cohort. This produces a more significant selective effect from frailty. The variance $\frac{k}{(\lambda(x))^2}$ is also decreasing with age. But the coefficient of variation for a Gamma distribution $\sqrt{\frac{1}{k}}$ is constant and does not change with age. This is the unique property of a Gamma distributed frailty, whereas other assumed forms of frailty usually exhibit a decreasing coefficient of variation. An example of this case is the Inverse Gaussian distributed frailty discussed next.

The distribution of frailty for the deaths at age $x$ is also Gamma with parameters $Gamma(k + 1, \lambda(x))$. To show this, the joint density of age and frailty derived in (8) is integrated over all possible values of $z$ to obtain the unconditional distribution for age $x$:

$$f_X(x) = \int_0^\infty \frac{\lambda^k}{\Gamma(k)} \cdot z^k \cdot e^{-\lambda(x) \cdot z} \cdot \mu(x)dz$$

$$= \mu(x, 1) \cdot k \cdot \frac{\lambda^k}{(\lambda(x))^{k+1}} \cdot \int_0^\infty \frac{(\lambda(x))^{k+1}}{\Gamma(k+1)} \cdot z^{k+1-1} \cdot e^{-\lambda(x) \cdot z}dz \tag{11}$$

$$= \mu(x, 1) \cdot k \cdot \frac{\lambda^k}{(\lambda(x))^{k+1}} \tag{12}$$

since the integrand in (11) is a Gamma density.

Hence the conditional density of frailty is:

$$f_{Z|X}^*(z|X = x) = f_{Z|X}(z|X = x) = \frac{f_{X,Z}(x,z)}{f_X(x)}$$

$$= \frac{\frac{\lambda^k}{\Gamma(k)} \cdot z^k \cdot e^{-\lambda(x) \cdot z} \cdot \mu(x, 1)}{\mu(x) \cdot k \cdot \frac{\lambda^k}{(\lambda(x))^{k+1}}}$$

$$= \frac{(\lambda(x))^{k+1}}{\Gamma(k+1)} \cdot z^{(k+1)-1} \cdot e^{-\lambda(x) \cdot z} \tag{13}$$

5

which is the density function of a $Gamma(k + 1, \lambda(x))$. The mean frailty for the deaths at age x is $\frac{k+1}{\lambda(x)}$, which is higher than the average frailty for survivors. Again this shows that, as the frailer individuals die first, the average frailty is decreasing as age increases.

A property of the Gamma distribution is that it can be scaled to obtain another Gamma distribution with the same shape parameter. If $X \sim Gamma(k, \lambda)$, then $\alpha X \sim Gamma(k, \frac{\lambda}{\alpha})$. If the distribution of frailty $z$ among survivors at age x is $Gamma(k, \lambda(x))$, then since $\mu(x, z) = z \cdot \mu(x, 1)$, the distribution for the force of mortality at age $x$ is $Gamma(k, \frac{\lambda(x)}{\mu(x,1)})$. Therefore:

$$f(\mu_x) = \frac{(\frac{\lambda(x)}{\mu(x,1)})^k}{\Gamma(k)} \cdot (\mu_x)^{k-1} \cdot e^{-\frac{\lambda(x)}{\mu(x,1)} \cdot \mu_x} \tag{14}$$

## 2.2 Inverse Gaussian distributed frailty

Another common form of distribution assumed for frailty is the Inverse Gaussian distribution. Under the Inverse Gaussian distributed frailty, the density function for frailty (Hougaard, 1984) is:

$$f_Z(z) = f_{Z|X}(z|X = 0) = (\frac{\delta}{\pi})^{\frac{1}{2}} \cdot e^{\sqrt{4\delta\theta}} \cdot z^{-\frac{3}{2}} \cdot e^{-\theta z - \frac{\delta}{z}} \tag{15}$$

The mean and variance of frailty at age 0 are:

$$E[z] = \bar{z}_0 = (\frac{\delta}{\theta})^{\frac{1}{2}}, \quad Var[z] = \frac{1}{2}\sqrt{\frac{\delta}{\theta^3}} \tag{16}$$

Similar to the Gamma distributed frailty, under the Inverse Gaussian distribution ($IG(\delta, \theta)$), the distribution of frailty for survivors at age x is also Inverse Gaussian ($IG(\delta, \theta(x))$), with:

$$\theta(x) = \theta + H(x)$$
$$= \theta + \int_0^x \mu(t)dt \tag{17}$$

The proof is similar to that for the Gamma distribution. The conditional distribution of frailty given survival to age $x$ ($f_x(z)$) is derived by integrating the joint distribution function of $x$ and $z$ with respect to age from $x$ to infinity:

$$f_{Z|X}(z|X = x) = f_{X,Z}(x, z|X > x) = \int_x^\infty f_{T,Z}(t, z)dt$$
$$= \int_x^\infty f_Z(z) \cdot f_{T|Z}(t|z)dt$$
$$= f_Z(z) \cdot \int_x^\infty f_{T|Z}(t|z)dt$$
$$= (\frac{\delta}{\pi})^{\frac{1}{2}} \cdot e^{\sqrt{4\delta\theta}} \cdot z^{-\frac{3}{2}} \cdot e^{-\theta z - \frac{\delta}{z}} \cdot e^{-z \cdot H(x)}$$
$$= (\frac{\delta}{\pi})^{\frac{1}{2}} \cdot e^{\sqrt{4\delta\theta}} \cdot z^{-\frac{3}{2}} \cdot e^{-(\theta + H(x))z - \frac{\delta}{z}} \tag{18}$$

The result is normalized to form another Inverse Gaussian density with

$$\delta' = \delta, \quad \theta' = \theta + H(x) \tag{19}$$

Therefore, $f_{Z|X}(z|X = x)$ is:

$$f_{Z|X}(z|X = x) = \left(\frac{\delta}{\pi}\right)^{\frac{1}{2}} \cdot e^{\sqrt{4\delta\theta(x)}} \cdot z^{-\frac{3}{2}} \cdot e^{-\theta(x)\cdot z - \frac{\delta}{z}} \tag{20}$$

The mean frailty of survivors up to age x under the Inverse Gaussian distributed frailty is $\left(\frac{\delta}{\theta(x)}\right)^{\frac{1}{2}}$, which is decreasing with age. The speed of decrease is higher if the population is more heterogeneous (smaller $\theta$) as is the variance $\frac{1}{2}\sqrt{\frac{\delta}{(\theta(x))^3}}$.

In contrast to the Gamma distributed frailty, the coefficient of variation $(4\delta\theta(x))^{-\frac{1}{4}}$ is a decreasing function of age, whereas under the Gamma assumption, the coefficient of variation is constant. The Inverse Gaussian distribution is also closed under scaling. $X \sim IG(\delta, \theta)$ can be scaled by $\alpha$ to form another Inverse Gaussian distribution: $\alpha X \sim IG(\alpha\delta, \frac{\theta}{\alpha})$. Therefore, if $z$ is $IG(\delta, \theta(x))$, the distribution for $\mu$, which is $z$ scaled by the standard force of mortality $\mu(x, 1)$, follows an $IG(\mu(x,1)\delta, \frac{\theta(x)}{\mu(x,1)})$:

$$f(\mu_x) = \left(\frac{\mu(x,1)\delta}{\pi}\right)^{\frac{1}{2}} \cdot e^{\sqrt{4\delta\theta(x)}} \cdot \mu_x^{-\frac{3}{2}} \cdot e^{-\frac{\theta(x)}{\mu(x,1)}\cdot\mu_x - \frac{\mu(x,1)\delta}{z}} \tag{21}$$

## 2.3 Parameter Estimation for Frailty Models

A commonly used approach to estimate the parameters for the standard force of mortality and frailty distribution based on observed population mortality data is the mean frailty approach (Vaupel et al. (1986), Butt and Haberman (2002)). Under the mean frailty approach, it is assumed that the observed force of mortality is the population average force of mortality of the cohort:

$$\bar{\mu}_x = \mu(x, 1) \cdot \bar{z}_x$$

and the number of deaths of the cohort $d_x$ follows a Poisson distribution with $\lambda = \bar{\mu}_x E_x$ assuming the cohort exposure to risk $E_x$ is large.

Under the mean frailty approach, there are in theory many different functional forms for the observed force of mortality and the distribution of frailty consistent with the population data. Both $\mu(x, 1)$ and $\bar{z}_x$ are difficult to separately identify (Elber and Ridder, 1982). The fitted models under the mean frailty approach are dependent on the choice of the standard force of mortality. Frailty provides the link between the standard force of mortality and the observed cohort force of mortality. These are not uniquely determined by the model estimation. In the case where the form for the standard force of mortality has a similar pattern to the observed cohort force of mortality, heterogeneity in the frailty distribution will appear to be insignificant.

In order to improve the estimation, an alternative method is proposed that takes into account the variability of the observed population data. Under a frailty model, the population is assumed to be heterogeneous, and the degree of heterogeneity is represented by the distribution of the unobserved frailty factor $z$. If the distributional form for the frailty factor (Gamma or Inverse Gaussian) is given, then the distribution for the individual force of mortality is known (scaled Gamma or scaled Inverse Gaussian).

The observed cohort deaths data is treated as a sample drawn from the population with size $E_x$. Since only the total number of deaths of the cohort can be observed, the only information available about the force of mortality is the observed cohort force of mortality estimated from $\frac{d_x}{E_x}$, which is the mean mortality rate of the sample.

The individual forces of mortality are randomly distributed with mean $E[\mu]$ and variance $Var[\mu]$. From the central limit theorem, the sample mean with sample size $E_x$ is approximately normally distributed with mean, $E[\mu]$, and variance $\frac{Var[\mu]}{E_x}$. The population mortality data likelihood is then determined based on this assumed distribution of the sample mean of $\mu$. The marginal distribution of frailty is assumed to follow a Gamma or Inverse Gaussian distribution.

The mean and variance of the individual force of mortality at age x under these two distributions is:

- Gamma distribution

$$E[\mu_x] = \frac{\mu(x) \cdot k}{k + H(x)}, \quad Var[\mu_x] = \frac{(\mu(x))^2 k}{(k + H(x))^2} \tag{22}$$

- Inverse Gaussian distribution

$$E[\mu_x] = \mu(x) \cdot (\frac{\delta}{\delta + H(x)})^{\frac{1}{2}}, \quad Var[\mu_x] = \frac{(\mu(x))^2}{2} \sqrt{\frac{\delta}{(\delta + H(x))^3}} \tag{23}$$

Therefore, the mean and variance for the sample mean $\hat{\mu}_x$ is:

- Gamma distribution

$$E[\hat{\mu}_x] = \frac{E_x \cdot E[\mu_x]}{E_x} = E[\mu_x] = \frac{\mu(x) \cdot k}{k + H(x)}$$

$$Var[\hat{\mu}_x] = \frac{E_x \cdot Var[\mu_x]}{(E_x)^2} = \frac{Var[\mu_x]}{E_x} = \frac{(\mu(x))^2 \cdot k}{E_x \cdot (k + H(x))^2} \tag{24}$$

- Inverse Gaussian distribution

$$E[\hat{\mu}_x] = \frac{E_x \cdot E[\mu_x]}{E_x} = E[\mu_x] = \mu(x) \cdot (\frac{\delta}{\delta + H(x)})^{\frac{1}{2}}$$

$$Var[\hat{\mu}_x] = \frac{E_x \cdot Var[\mu_x]}{(E_x)^2} = \frac{Var[\mu_x]}{E_x} = \frac{(\mu(x))^2}{2 \cdot E_x} \sqrt{\frac{\delta}{(\delta + H(x))^3}} \tag{25}$$

The normal distribution can be fully specified by the first 2 moments (Wackerly, 2008), and

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x - \mu)^2}{2\sigma^2}) \tag{26}$$

where $\mu$ and $\sigma$ are the mean and standard deviation of the normal distribution.

The log likelihood function of the observed cohort force of mortality to be maximized is a function of $k(\delta)$, and the standard force of mortality parameters:

$$L(\hat{\mu}_x | E_x, k(\delta), \mu(x)) = \sum_{x,i} \left\{ -\frac{1}{2} [\log(2\pi) + \log(\sigma^2)] - \frac{(\hat{\mu}_x - \mu)^2}{2\sigma^2} \right\} \tag{27}$$

with $\mu$ and $\sigma^2$ being the respective mean and variance of the frailty distribution. The empirical distribution is used to estimate the standard force of mortality to minimize bias for the selected parametric form.

# 3  Markov Aging Model

Lin and Liu (2007) [7] developed and estimated a Markov aging model to describe the aging process of the human body. Studies in human physiology suggest that aging of human beings is associated with the change of a wide range of physiological functions, such as disturbances of metabolism and rarefaction of bone structure. Deterioration of physiological functions can be viewed as the worsening of health status, and as the body becomes less functional, individuals are more subject to disease and death. The concept of physiological age is introduced. The physiological age of an individual represents the degree of aging in the human body, and each physiological age represents a different level of functionality of the human body. Change in physiological age represents the decline in human body function, and the aging process is modelled in terms of a physiological age. Most functional variables reach their maximum between age 3 to 20, and then start declining roughly linearly, although individuals are heterogeneous in terms of the rate of decline.

The Markov model is specified based on the results of these studies. The Markov aging model is a continuous-time discrete-state multi-state model with states defined by physiological ages and death. Since human aging is irreversible, the transition between states is assumed to be one-directional.

At state $i$, an individual can move either to the next state or to the death state, which is an absorbing state. The transition rate matrix with $n$ physiological ages (total of $n+1$ states) is given by:

$$\Sigma = \begin{pmatrix} -(\lambda_1 + q_1) & \lambda_1 & 0 & \cdots & 0 & q_1 \\ 0 & -(\lambda_2 + q_2) & \lambda_2 & \cdots & 0 & q_2 \\ 0 & 0 & -(\lambda_3 + q_3) & \cdots & 0 & q_3 \\ \hdotsfor{6} \\ 0 & 0 & 0 & \cdots & -q_n & q_n \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

where $\lambda_i$ represents the rate of transition from state $i$ into the next state, and $q_i$ represents the rate of transition into the absorbing death state.

Since many physiological functions exhibit linear decline after a certain age, it is assumed that the transition rate $\lambda_i$ is constant after physiological age $k$ so that:

$$\lambda_i = \lambda, \quad \text{for } i > k \tag{28}$$

The death rate $q_i$ in different states varies to reflect the mortality risk due to different health conditions, and is assumed to be an increasing function of the number of the state $i$ after the initial $k$ developmental periods. The Australian population mortality rate has an approximate exponential growth at older ages. The following assumption for the death rate is used:

$$q_i = \gamma + \alpha e^{\beta i}, \quad \text{for } i > k \tag{29}$$

with $\gamma$ a health-independent background rate, allowed to be different between states to capture the mid age hump in observed mortality data. $\alpha e^{\beta i}$ is a health-dependent component. In developmental period $k$ each state has unique rates of transition $\lambda$ and $q$.

The transition rate matrix for the transient states is then:

$$\Lambda = \begin{pmatrix} -(\lambda_1 + q_1) & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & -(\lambda_k + q_k) & \lambda_k & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -(\lambda + \gamma + \alpha e^{\beta(k+1)}) & \lambda & \cdots & 0 \\ 0 & \cdots & 0 & 0 & -(\lambda + \gamma + \alpha e^{\beta(k+2)}) & \cdots & 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -(\alpha + e^{\beta n}) \end{pmatrix}$$

For these model assumptions, the time to death follows a phase-type distribution with $n + 1$ phases. See Neuts (1981) for details of phase-type distribution. Under a phase-type distribution, the survival function for time to death $X = x$ has a simple expression:

$$s(x) = \alpha \exp(\Lambda x)\mathbf{e} \tag{30}$$

where $\alpha$ is the initial state vector. At age 0, it is a vector of zeros with the first entry unity. $\exp(\Lambda t)$ is the matrix exponential of the product of time and the transition rate matrix for transient states. $\mathbf{e}$ is a vector of ones. The death probability under the phase-type distribution is then:

$$\hat{q}_x = \frac{s(x) - s(x+1)}{s(x)} \tag{31}$$

The parameters are estimated by minimizing the weighted sum of squared errors for the death probability $q_x$:

$$\text{SSE}(q_x | s_x, \lambda_i, q_i, \lambda, \gamma, \alpha, \beta) = \sum_{x=0}^{\omega-1} (q_x - \hat{q}_x)^2 \cdot s_x \tag{32}$$

where $s_x$ is the observed survival probabilities. These are used as the weight factor so that the squared errors at later ages, when $q_x$ is larger, are given reduced weight, since $s_x$ and its variability decrease with age. The parameters to be estimated are $\lambda_i$, $q_i$ (for $0 < i < k$), $\lambda$, $\gamma$, $\alpha$, and $\beta$.

## 4 Data

Mortality data for Australia is obtained from the Human Mortality Database. Cohort data for the 1940 and 1945 birth cohorts are used for both models, since they represent the recently retired population with the greatest potential demand for annuity contracts.

For the frailty model, the required format of the data is the cohort central exposure to risk, and cohort force of mortality. The cohort central exposure to risk is directly available from HMD. The observed cohort force of mortality is estimated by the central rate of death $m_x$, also available from the HMD. Since the Gompertz law is assumed for the standard force of mortality, which is an increasing function of age and is only suitable for adult mortality, the age range of 30 onwards is selected. The cohort force of mortality for birth cohort 1940 and 1945 for both males and females is shown in Figure 1. On the log scale these are close to linear and support the use of the Gompertz force of mortality assumptions for the age range considered.

## Figure 1: Observed Cohort Force of Mortality: Log Transform
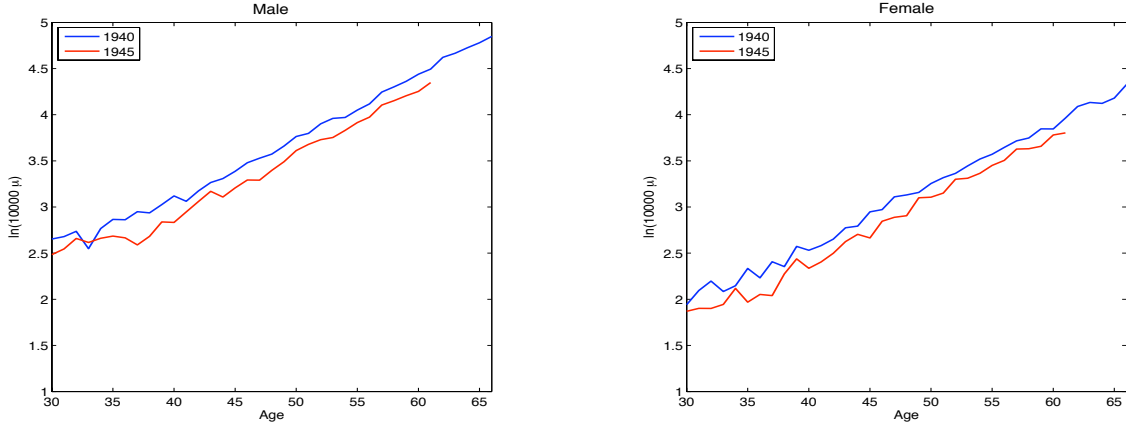


For the Markov Aging Model, the required format of data is the cohort death probability $q_x$ and survival probability $s_x$ at different ages for cohorts. $q_x$ is not directly available on HMD, so it is estimated from the central death rate $m_x$, which is available, by assuming a uniform distribution of death (UDD) during the year:

$$q_x = \frac{m_x}{1 + \frac{1}{2}m_x} \tag{33}$$

$s_x$ is then obtained from the calculated value of $q_x$:

$$s_x = \prod_{k=0}^{x}(1 - q_k) \tag{34}$$

The death probability $q_x$ (log scale) for the birth cohorts 1940 and 1945, both males and females, is shown in Figure 2.

The Markov model is used to fit the full age range and the non-linear form for the younger ages can be more flexibly handled with this model compared to the frailty model.

## Figure 2: Observed Cohort Death Probability: Log Transform



11
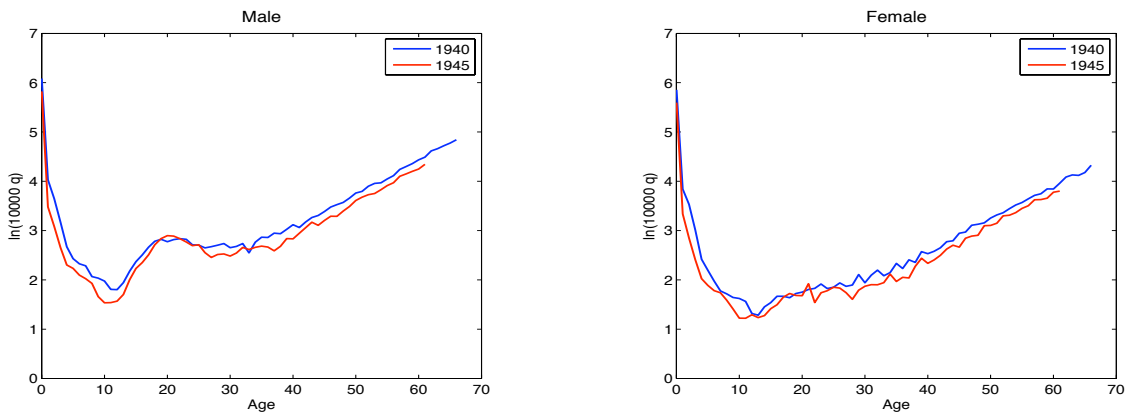
# 5 Results

## 5.1 Model Estimation

### 5.1.1 Frailty Model

The estimated maximum likelihood values for the frailty distribution and the standard force of mortality parameters are shown in Table 1.

Table 1: Estimated Parameters for Frailty Model

|  | 1940 Male | 1945 Male | 1940 Female | 1945 Female |
|---|---|---|---|---|
| Gamma |  |  |  |  |
| Frailty parameter | 0.21108 | 0.07775 | 0.16847 | 0.07170 |
| $\alpha$ | 0.00012 | 0.00008 | 0.00007 | 0.00005 |
| $\beta$ | 0.08436 | 0.09981 | 0.08052 | 0.09218 |
| Maximum Likelihood | -4927.22546 | -3360.19872 | -1223.90347 | -903.08733 |
| Inverse Gaussian |  |  |  |  |
| Frailty parameter | 0.00005 | 0.00003 | 0.00009 | 0.00005 |
| $\alpha$ | 0.00189 | 0.00208 | 0.00063 | 0.00051 |
| $\beta$ | 0.14328 | 0.14701 | 0.13455 | 0.14178 |
| Maximum Likelihood | 276.05935 | 227.67831 | 274.60426 | 232.15072 |

For the frailty model, mortality heterogeneity for both male and female cohorts is significant, as indicated by the small value of the frailty parameter. The estimated average force of mortality (log transform) of the cohort is plotted and compared with the observed cohort force of mortality in Figure 3.

Frailty is unobserved and there is no biological reason as to which distribution should be selected for the frailty distribution. The Inverse Gaussian distribution provides a better fit to observed data and is selected. Figure 4 shows the projection of cohort average force of mortality (log transform) to the higher ages for the Inverse Gaussian assumption.

### 5.1.2 Markov Aging Model

For the Markov ageing model estimated parameters are given in Table 2. Figure 5 shows the fitted death probability, the observed death probability (log scale), and predictions for higher ages. An important difference between the frailty model and the Markov ageing model can be seen from these plots. For the frailty model the assumption of Gompertz mortality leads to a linear projection of future mortality rates at the older ages. For the Markov model the model forecasts a decline in mortality rates at the older ages (on the log scale).

The model provides a good fit for all 4 cohorts given the number of parameters involved. There are 12 for male or 15 for female as opposed to over 100 parameters in models such as the Lee-Carter model (Lin and Liu, 2007)). To analyze the goodness-of-fit of the model, the $R^2$ coefficient is calculated for the cohorts. $R^2$, the coefficient of determination, is the proportion of variation in the observed data that is explained by the model. In the Markov aging model, the total variation in observed data is defined

Figure 3: Observed v.s. Fitted Cohort Average Force of Mortality: Frailty Model
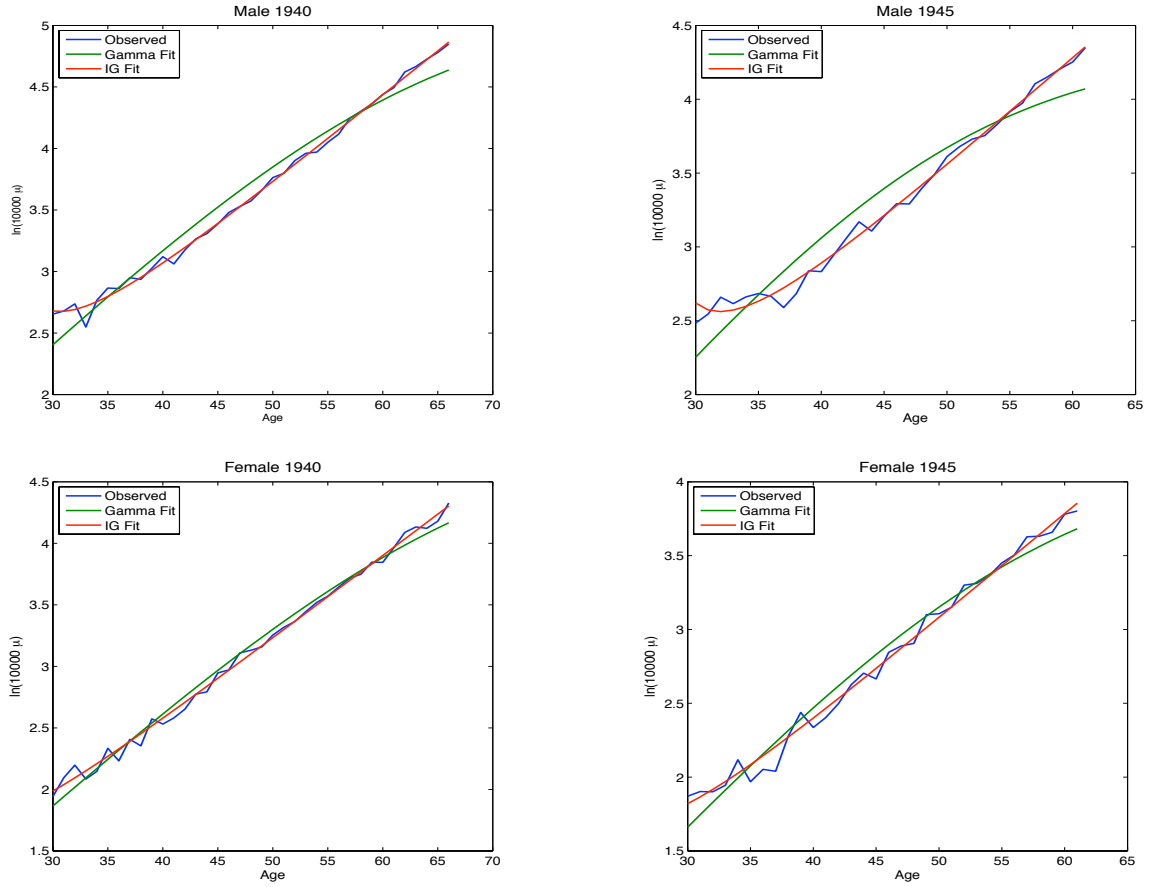
as:

$$\text{SST} = \sum_{x=0}^{\omega-1} (q_x - \bar{q})^2 \cdot s(x) \tag{35}$$

where $\omega$ is the observed highest age, and $\bar{q}$ is the average death probability at all observed ages. The variation that is not explained by the model is the weighted sum of squared errors:

$$\text{SSE} = \sum_{x=0}^{\omega-1} (q_x - \hat{q}_x)^2 \cdot s(x) \tag{36}$$

where $\hat{q}_x$ is the model fitted death probability. The proportion of variation that is explained by the model is:

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} \tag{37}$$

The $R^2$ coefficients for the Markov ageing model are shown in Table 2. The $R^2$ for these cohorts indicate a satisfactory fit.

## 5.2 Mortality Heterogeneity

### 5.2.1 Frailty Model

For the frailty model, the heterogeneity of the population is determined by the distribution of frailty factor. The more disperse the distribution is, the more heterogeneous

13

Figure 4: Predicted Cohort Force of Mortality at Higher Ages: Frailty Model
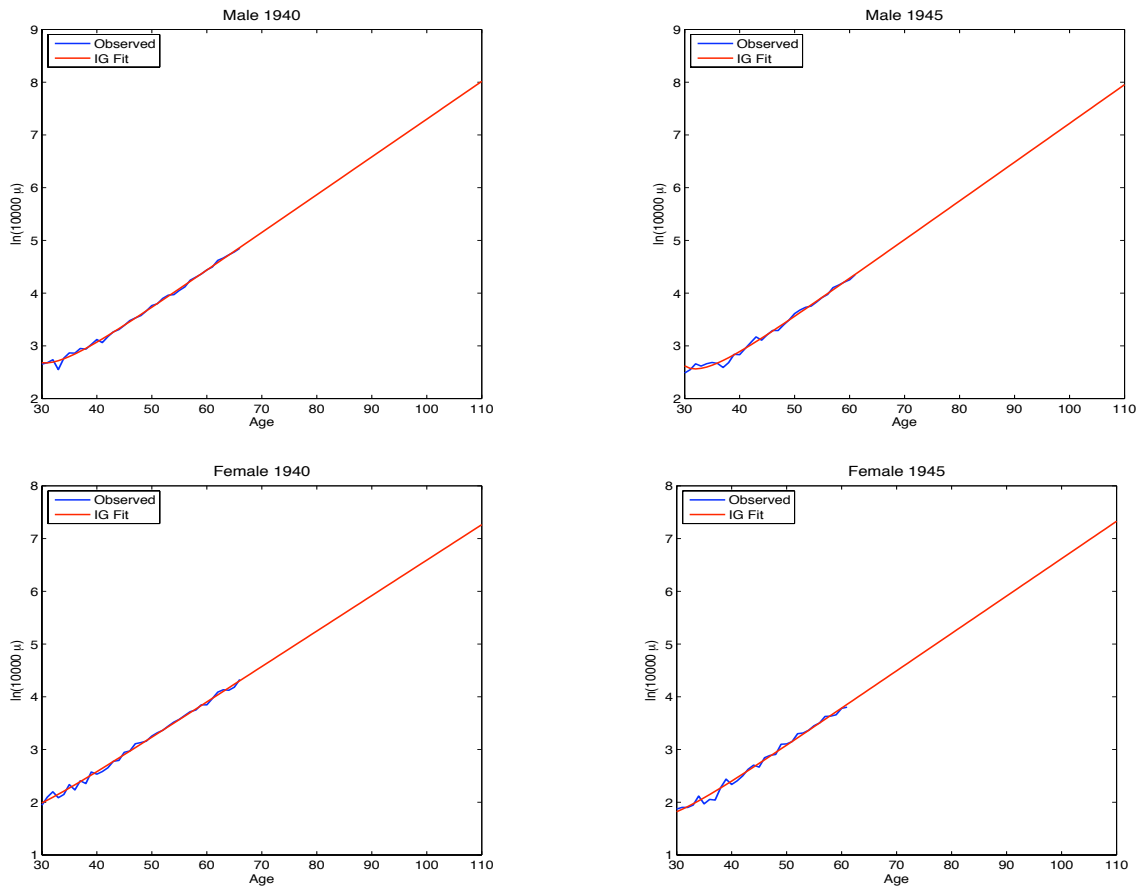


the population. The probability density function of frailty is shown in Figure 6. The distribution of frailty is heavily positively skewed for both males and females. At age 0, the majority of the population is concentrated at a low level of frailty with a long tail to the right with mean frailty at age 0 equal to 1. As age increases, the more frail individuals have a much higher chance of dying, and contribute to the majority of deaths at early ages. As the more frail individuals die out, the survivors are more concentrated to the left in the frailty distribution, and the selective effect of frailty results in the remaining cohort having a much lower mean frailty. From age 0 to age 30, the change of shape of the density curve in the plots is not significant. From age 30 onwards, the density curve shrinks to the left, and at age 90, the majority of survivors have a frailty value very close to 0.

The mean frailty of the cohort at each age is shown in Table 3. At age 0, a mean frailty of 1 is assumed. At age 30, the mean frailty drops significantly for the 1940 male cohort due to the deaths of the high frailty individuals. At age 90, the average frailty is very low, and the majority of survivors are concentrated in the low frailty range.

The standard deviation of frailty at different ages is shown in Table 5. The heavy skewness of the frailty distribution results in an extremely high standard deviation of frailty at age 0. As age increases the standard deviation of frailty reduces significantly.

Mortality rates for individuals with different levels of frailty are also compared.

14

Figure 5: Observed v.s. Fitted Death Probability: Markov Aging Model
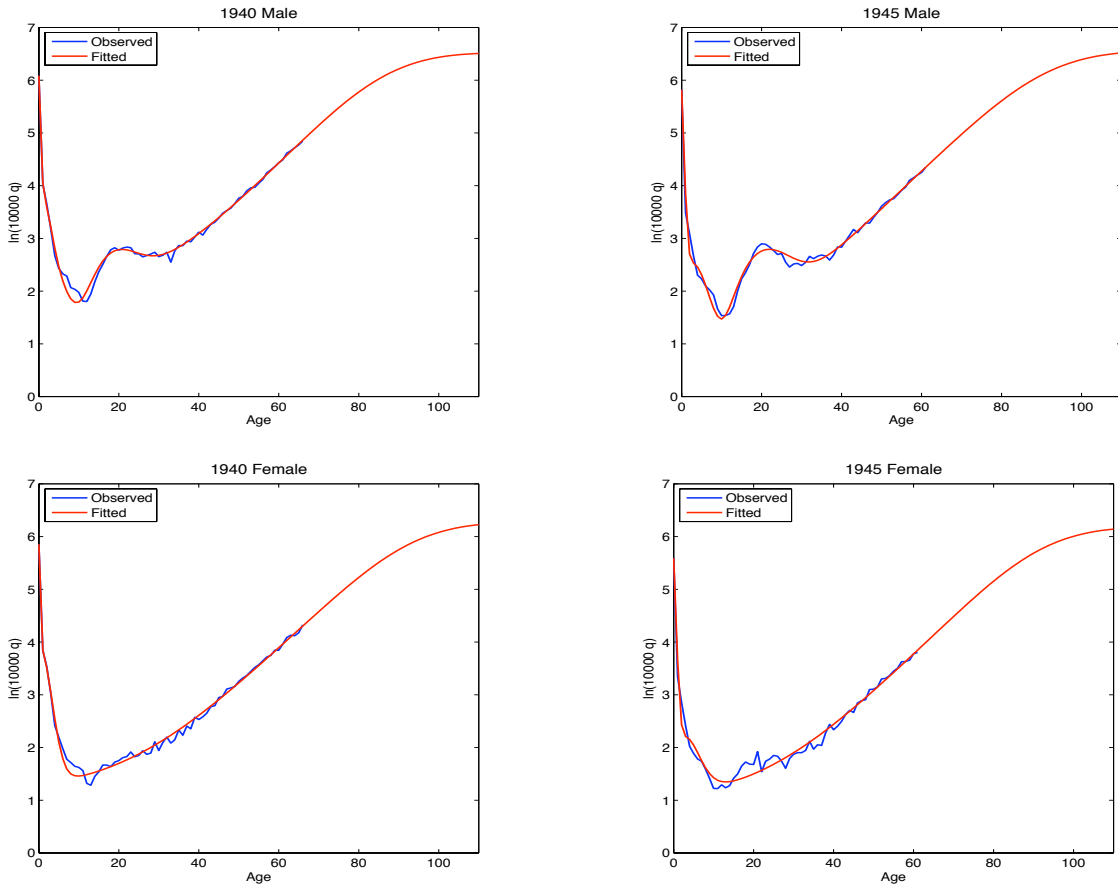
The death probability $q(x,z)$ is estimated from the individual force of mortality using:

$$q(x,z) = 1 - e^{-\mu(x,z)} \tag{38}$$

with the individual force of mortality $\mu(x,z)$ assumed constant through each year age interval. The mortality rates from the frailty model for individuals with frailty 1, 0.01, 0.001, 0.0005, 0.0001, and 0.00005 are shown in Figure 7.

The plots show that the aging of the cohort as a whole is much slower than that for individuals. The slope of the curve for the cohort is much lower, especially at high ages where the individual mortality rates curves become substantially steeper. For a specific individual, if the individual is healthy with a low frailty, the chance of dying will stay low even as age increases. Increasing frailty from higher susceptibility to disease, which would correspond to a higher frailty, increases the chance of dying significantly. The mortality rates for an individual vary significantly as shown in Figure 7. Heterogeneity is significant with substantial differences in survival prospects for individuals with differing frailties.

### 5.2.2 Markov Aging Model

For the Markov aging model, the heterogeneity of population mortality is measured by the distribution of physiological ages through time. Heterogeneity reflects the different health conditions of individuals at the same age. Under the phase-type distribution,

## Table 2: Estimated Parameters for Markov Aging Model

|  |  | 1940 Male | 1945 Male | 1940 Female | 1945 Female |
|---|---|---|---|---|---|
| General Parameters | $\lambda$ | 1.1018615 | 1.0635349 | 1.0152901 | 1.0439359 |
|  | $\alpha$ | 0.0000544 | 0.0000750 | 0.0000445 | 0.0000396 |
|  | $\beta$ | 0.0715661 | 0.0688020 | 0.0710774 | 0.0713141 |
|  | $\gamma$ | 0.0006377 | 0.0001717 | 0.0003111 | 0.0002565 |
| Developmental Period | $\lambda_1$ | 3.1885624 | 2.0523974 | 5.7001226 | 2.0078893 |
|  | $q_1$ | 0.1457796 | 0.0812789 | 0.1858058 | 0.0633914 |
|  | $\lambda_2$ | 0.7862403 | 0.6058546 | 1.0377221 | 0.6080937 |
|  | $q_2$ | 0.0000000 | 0.0000000 | 0.0027803 | 0.0000000 |
|  | $\lambda_3$ | 0.8462262 | 0.6977650 | 1.0157749 | 0.6996102 |
|  | $q_3$ | 0.0130941 | 0.0000000 | 0.0085292 | 0.0000000 |
|  | $\lambda_4$ | 0.8476995 | 0.7001925 | 1.0181996 | 0.6996077 |
|  | $q_4$ | 0.0000000 | 0.0044590 | 0.0039534 | 0.0030335 |
| Special Background Rates | Period 2 | (11,17) | (11,16) | N/A | N/A |
|  | $\gamma_2$ | 0.0002056 | -0.0000635 | N/A | N/A |
|  | Period 3 | (18,27) | (17,28) | N/A | N/A |
|  | $\gamma_3$ | 0.0016857 | 0.0016057 | N/A | N/A |
| Weighted Least Square |  | 0.0000009 | 0.0000031 | 0.0000007 | 0.0000019 |

## Table 3: Goodness-of-Fit: $R^2$ Coefficients

| Cohort | 1940 Male | 1945 Male | 1940 Female | 1945 Female |
|---|---|---|---|---|
| $R^2$ | 0.9995 | 0.9972 | 0.9994 | 0.9973 |

assuming the initial state is 1, the probability for an individual aged $x$ to be in state $i$ (denoted by $P_i(x)$) is given by the $i$-th entry of the vector $[\boldsymbol{\alpha} \exp(\boldsymbol{\Lambda}x)]$:

$$P_i(x) = Pr(I = i, X = x) = [\boldsymbol{\alpha} \exp(\boldsymbol{\Lambda}x)]_i \tag{39}$$

The conditional probability of being in state $i$, given surviving to age x is:

$$\pi_i(x) = \frac{P_i(x)}{s(x)} = \left[\frac{\boldsymbol{\alpha} \exp(\boldsymbol{\Lambda}x)}{\boldsymbol{\alpha} \exp(\boldsymbol{\Lambda}x)\mathbf{e}}\right]_i \tag{40}$$

Therefore, $\pi(x)$ is the empirical density function for the distribution of physiological age at age x. The plotted density curve for the distribution of physiological age at different ages is shown in Figure 8.

## Table 4: Mean Frailty of Cohort at Different Ages

| Age | 1940 Male | 1945 Male | 1940 Female | 1945 Female |
|---|---|---|---|---|
| 0 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| 35 | 0.00575 | 0.00389 | 0.01374 | 0.01114 |
| 50 | 0.00171 | 0.00108 | 0.00479 | 0.00360 |
| 65 | 0.00057 | 0.00035 | 0.00174 | 0.00123 |
| 80 | 0.00020 | 0.00012 | 0.00063 | 0.00043 |

Figure 6: Distribution of Frailty at different ages
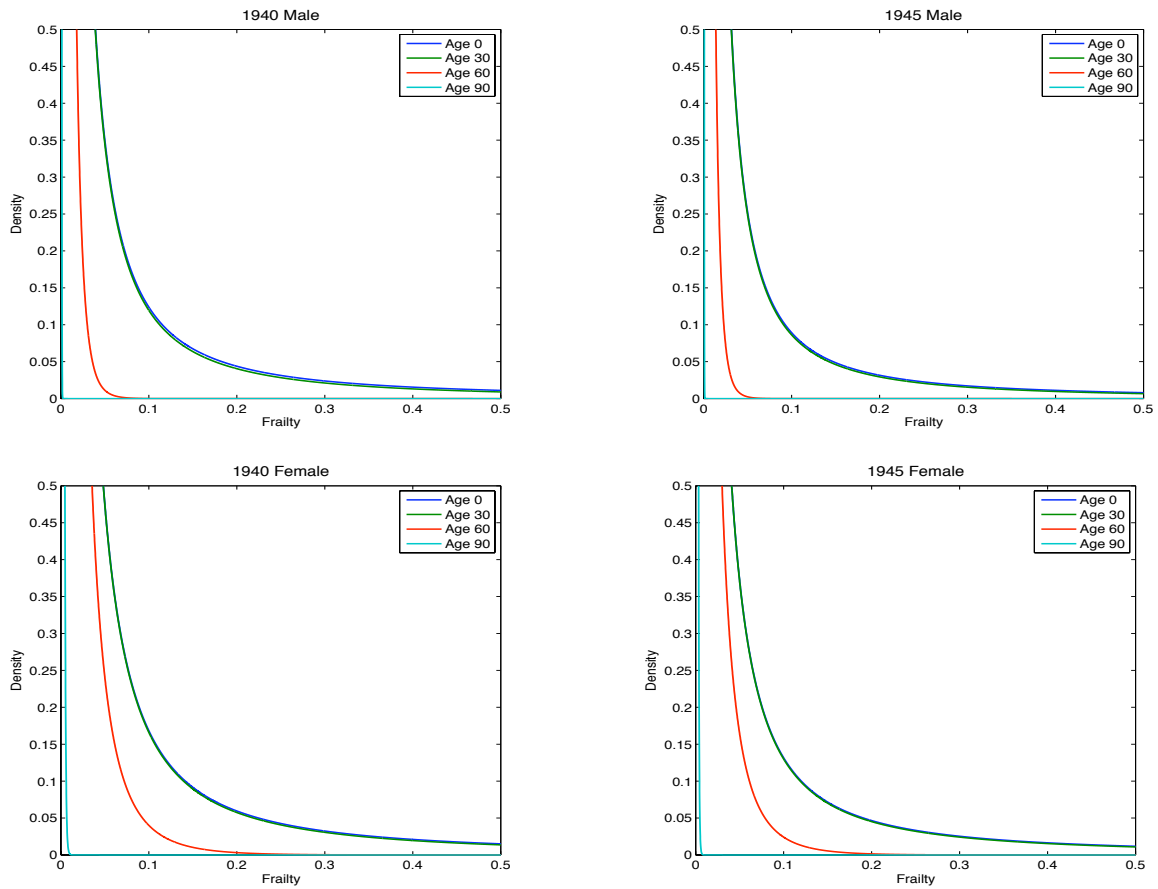


The model implicitly assumes that initially all individuals are physiological age 1. Thereafter the aging patterns of individuals are allowed to differ. At lower ages, the distribution is more concentrated, with the cohort at lower ages less heterogeneous. As age increases, the density curve flattens, and the level of heterogeneity of the cohort increases with age.
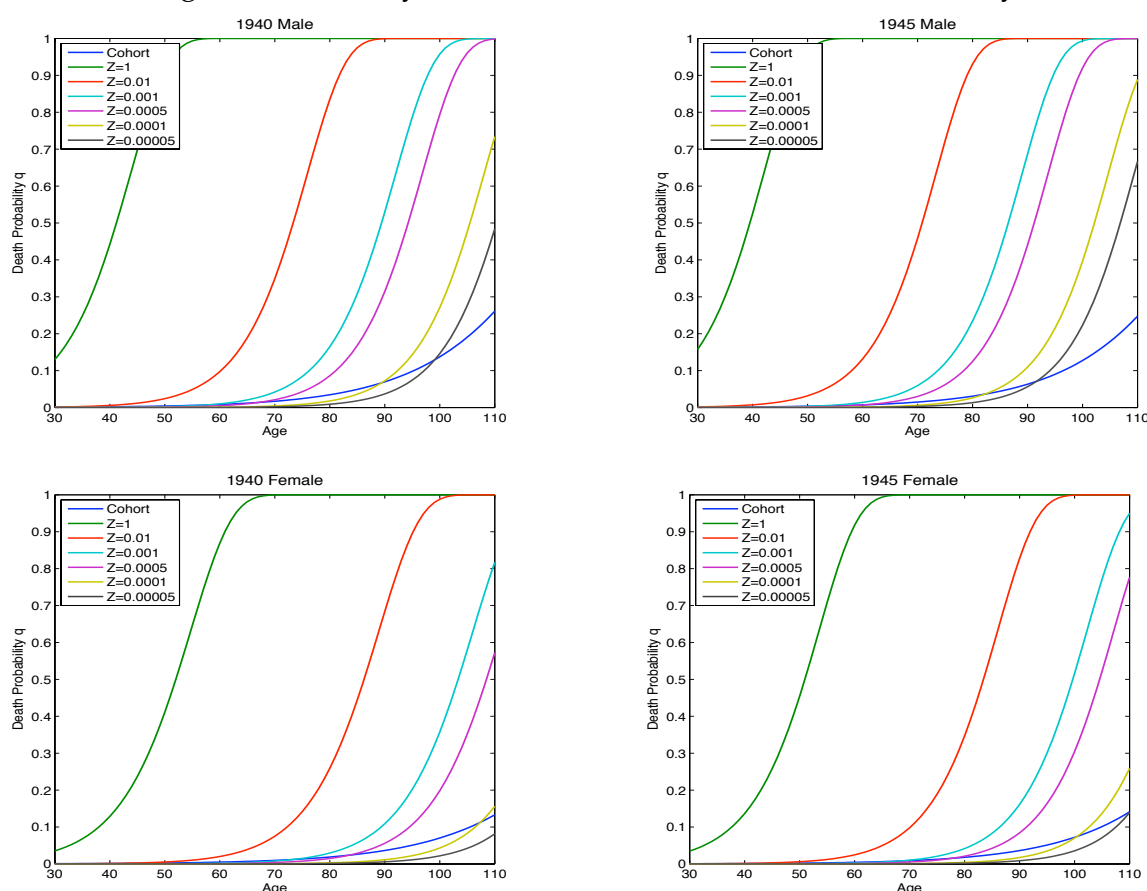
## 5.3 Implications for Annuity Market

Both models for heterogeneity have implications for annuity markets. If the heterogeneity is not significant then annuity rates will not vary much for any age and will be close to the cohort annuity rates. However if annuity rates are found to vary significantly for individuals with different levels of mortality based on the model results then

Table 5: Standard Deviation of Frailty at Different Ages

| Age | 1940 Male | 1945 Male | 1940 Female | 1945 Female |
|-----|-----------|-----------|-------------|-------------|
| 0   | 101.80271 | 141.31027 | 74.89512    | 95.77401    |
| 35  | 0.04439   | 0.03428   | 0.12059     | 0.11261     |
| 50  | 0.00718   | 0.00505   | 0.02483     | 0.02071     |
| 65  | 0.00140   | 0.00094   | 0.00542     | 0.00416     |
| 80  | 0.00028   | 0.00018   | 0.00119     | 0.00084     |

Figure 7: Mortality Rates of Individuals with Different Frailty

this has significant implications for pricing and underwriting of life annuities. In order to consider the implications for the life annuity market, annuity rates are computed using the estimated models and projected future mortality for the cohorts.

Tables 6 and 7 show the annuity rates for a male individual assumed to be 65 under the two models for the 1940 male cohort. The life annuity contracts included are a whole life annuity at age 65 and a deferred whole life annuity with a deferred period of 20 years assuming an interest rate of 3%. The deferred whole life annuity has payments starting from age 85.

Table 6: Annuity Rate for Individuals with Different Frailty: 1940 Male

| Frailty | Cohort | 0.00005 | 0.0001 | 0.0002 | 0.0005 | 0.001 | 0.01 |
|---------|--------|---------|--------|--------|--------|-------|------|
| $q_{65}$ | 0.012 | 0.001 | 0.002 | 0.004 | 0.010 | 0.021 | 0.189 |
| Whole Life Annuity | $14.31 | $18.12 | $16.36 | $14.38 | $11.49 | $9.18 | $2.59 |
| Deferred Life Annuity | $2.36 | $4.32 | $2.94 | $1.66 | $0.47 | $0.08 | $0.00 |
| F(z) | | 19.40% | 38.26% | 56.95% | 76.49% | 86.70% | 99.59% |

The tables have been constructed so that annuity rates shown correspond to approximately the same proportion of the population for each case. These proportions are shown in the last row of each table and denoted by the $F(z)$ and $F(j)$. For the Markov model the physiological ages have a roughly equivalent mapping to the frailty factor for these proportions of individuals in the population.

18

Figure 8: Distribution of Physiological Age at Different Ages: Male
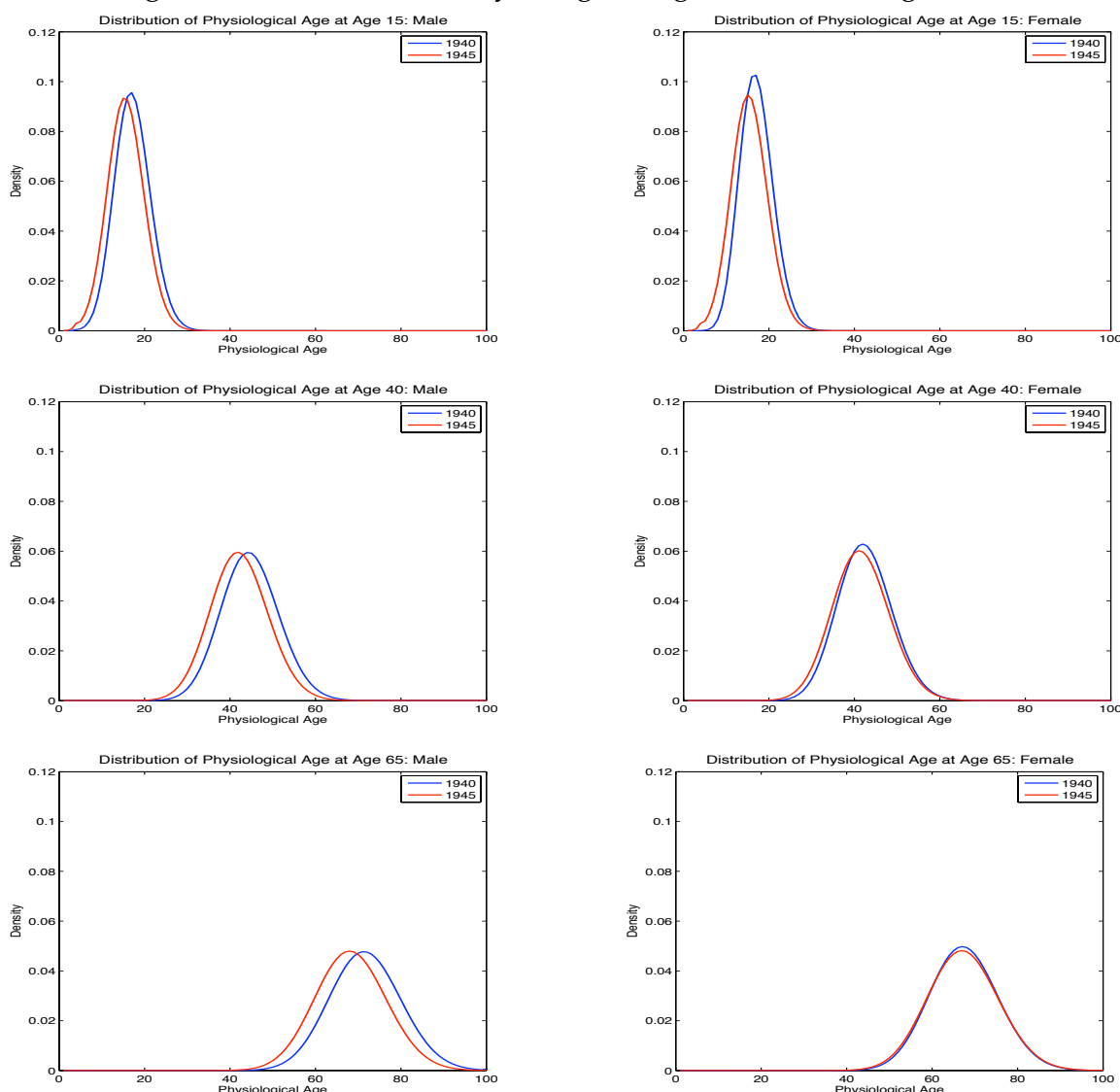
The life annuity rates decrease significantly as the health condition of an individual decreases, which is measured by the increase in frailty factor under the frailty model and increase in physiological age under the Markov Aging Model.

The frailty model has a wider range of annuity values. The Markov Aging model life annuity values are generally higher. For the frailty model, the healthiest 20% of the population would pay a purchase price of $18.12 for each $1 of annuity income whereas in the Markov model the healthiest 20% would pay $19.44 for every $1 of annuity income. Similarly for the deferred annuities. The healthiest 20% would pay $4.32 under the frailty model for every $1 of annuity income commencing at age 85 whereas the for the Markov model they would pay $4.55.

For the approximately least healthy 13% of individuals the frailty model produces a life annuity rate of $9.18 for every $1 of annuity income whereas the Markov model produces an annuity rate of $14.44.

The difference in results between the two models reflects the differing assumptions as to how mortality heterogeneity is measured. Frailty is a health factor fixed at birth. Given survival an individual's percentile in the cohort is increasing with age. In the

Table 7: Annuity Rates for Individuals with Different Physiological Age: 1940 Male

| Physiological Age j | 64 | 68 | 73 | 77 | 81 | 94 |
|---|---|---|---|---|---|---|
| $q_{65}$ | 0.006 | 0.008 | 0.011 | 0.015 | 0.019 | 0.047 |
| Whole Life | $19.44 | $18.31 | $16.83 | $15.63 | $14.44 | $11.15 |
| Deferred Life | $5.34 | $4.55 | $3.63 | $2.99 | $2.46 | $1.50 |
| F(j) | 19.47% | 35.49% | 59.01% | 75.81% | 87.83% | 99.60% |

Markov aging model, the distribution by physiological age is changing over time. As the whole cohort ages a surviving individual moves into a physiological age according to the estimated transition probabilities.

These models are based only on population level data. They imply distributions of individuals heterogeneity based on model assumptions calibrated to data. They do highlight the extent of heterogeneity. Given knowledge of an individual's relative health they allow the determination of an annuity rate that reflects their survival prospects. The results clearly indicate that there is substantial heterogeneity in the population.

If a life annuity market is to be made viable for a wider range of individuals other than the most healthy lives it will be essential to understand the major factors determining heterogeneity and to assess mortality in the underwriting process.

# 6 Conclusion

This paper has quantified heterogeneity in the Australian population mortality using the well known frailty models as well as a more recently developed Markov ageing model. Both models have their advantages and disadvantages. Neither model provides an explicit basis for incorporating heterogeneity in life annuity pricing but they do allow a quantification of the importance of heterogeneity in life annuity pricing.

The frailty model was found to be heavily dependent on the underlying assumptions. It is difficult to differentiate between the volatility in mortality rates arising from heterogeneity and the natural random variation in mortality rates through time. Despite this both models provide a guide to the expected variability required in life annuity rates to allow for heterogeneity.

The models can be used in pricing if it is possible to associate the different levels of mortality in the models with causal factors such as health status and socio-economic status. If pensions and annuities are to be provided to a broader population than the individuals who self select to purchase life annuities in the private annuity market, then the results show clearly that heterogeneity must be taken into account in annuity pricing.

# 7 Acknowledgement

# References

[1] Butt, Z., Haberman, S.. 2002. Application of Frailty-Based Mortality Models to Insurance Data. *Actuarial Research Paper No. 142, Department of Actuarial Science and Statistics, City University, London.*

[2] Elbers, C., Ridder, G., 1982. True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. *The Review of Economics Studies, Vol. 49, No. 3: 403-409.*

[3] Ganegoda, A., Bateman, G., 2008. Australias disappearing market for life annuities. *UNSW Centre for Pensions and Superannuation Discussion Paper 01/08*

[4] Hougaard, P., 1984. Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity. *Biometrika, Vol. 71, No. 1: 75:83.*

[5] Hougaard, P., 1986. Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika, 73:671-678.*

[6] Hoem, J. M., 1990. Identifiability in Hazard Models with Unobserved Heterogeneity: The Compatibility of Two Apparently Contradictory Results. *Theoretical Population Biology, 37: 124-128.*

[7] Lin, X. S., Liu, X.. 2007. Markov Aging Process and Phase-Type Law of Mortality. *North American Actuarial Journal Vol. 11, No. 4: 92-109.*

[8] Neuts, M. F., 1981. Matrix-Geometric Solutions in Stochastic Models. *Baltimore: Johns Hopkins University Press.*

[9] Olivieri, A.. 2006. Heterogeneity in Survival Models, Applications to Pensions and Life Annuities. *Belgian Actuarial Bulletin, 6: 23-39.*

[10] Vaupel, J. W., Manton, K. G., Stallard, E., 1979. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography Vol. 16, No. 3: 439-454.*

[11] Vaupel, J. W., Manton, K. G., Stallard, E., 1981. Methods for Comparing the Mortality Experience of Heterogeneous Populations. *Demography, Vol. 18, No. 3: 389-410.*

[12] Vaupel, J. W., Manton, K. G., Stallard, E.. 1986. Alternative Models for the Heterogeneity of Mortality Risks Among the Aged. *Journal of the American Statistical Asosciation Vol. 81 No. 395: 635-644 .*

[13] Wackerly, D. D., Mendenhall, W., Scheaffer, R. L., 2008. Methematical Statistics with Applications, Seventh Edition. *Thomson Press.*

[14] Zuev, S. M., Yashin, A. I., Manton, K. G., Dowd, E., Pogojev, I. B., Usmanov, R. N.. 2000. Vitality Index in Survival Modeling: How Physiological Aging Inuences Mortality. *Journal of Gerontology, 55A: 10-19.*