Annuities and Aggregate Mortality Uncertainty*

Justin van de Ven and Martin Weale
National Institute of Economic and Social Research,
2, Dean Trench Street,
London SW1P 3HE

10th March 2007

Abstract

This paper explores the effect of aggregate mortality risk on thepricing of annuities. It uses a two-period model; in the second period people face a constant but intiially unknown risk of death. Old people can either carry the aggregat emortlaity risk for themselves or buy annuities which are sold by young people. A market-clearing price for such annuties is established. It is found that old people would, given the choice, decide to carry a considerable part of aggregate mortality risk for themselves.

*This work has been supported by the ESRC under its World Economy and Finance Programme. Please treat the results presented here as preliminary.

1 Introduction

Many of the issues associated with annuity markets have been well-researched in recent years, as reported in the comprehensive survey by Davidoff, Brown, and Diamond (2005). Much of this research has focussed upon the merits of annuities, with the general finding that there is likely to be a substantial role for annuitisation even in context of heterogeneous preferences for investment, the time stream of consumption, and bequests. In contrast, very little work has been conducted to understand the workings of annuities markets with regard to aggregate mortality risk. This is surprising, given the role played by mortality risk in the *a priori* justification for annuities. Here, we suggest an appropriate economic model for pricing annuities, and use the model to project plausible values for the annuity risk premium attributable to uncertainty regarding individual and cohort specific life expectancy.

To clarify terms, aggregate mortality risk refers to the uncertainty that is associated with the average longevity of a birth cohort. Individual mortality risk, by contrast, is the uncertainty of the precise time of death of an individual, relative to their respective birth cohort. Individuals from a given birth cohort can protect themselves against individual mortality risk through the use of a tontine, which pays an annual dividend that depends upon the realised mortality rate of the cohort. Purchasers of such a tontine will, nevertheless remain exposed to income uncertainty to the extent that cohort mortality is uncertain. This is as opposed to a standard life annuity, which pays a guaranteed income to purchasers until death.[1]

The topic of aggregate mortality uncertainty was not discussed in the survey by Davidoff, Brown, and Diamond (2005). While Bohn (2005) does air the question of uncertain longevity, he limits himself to scenario analysis – the effect of a permanent increase in longevity during retirement – and does not address the question of risk in its more conventional sense. The problem is that the true mortality rate of any particular cohort is known only *ex post*. When an insurance company sells an annuity it contracts to pay a regular income calculated on the basis of a forecast of the mortality rate. Like any forecast, projections of mortality rates are subject to error, and any insurance company selling annuities therefore carries mortality risk. Casual discussion in the United Kingdom has suggested that this risk may be substantial. Banks and Blundell (2005) make the point that "unanticipated increases in longevity put pressure on all forms of pension systems" although they

[1] We are not aware of any contemporary sellers of such tontines in the United Kingdom but it is possible that there would be benefits to filling this gap in financial markets; certainly, if the premium associated with aggregate mortality risk is large, then one might expect that annuitants would value the opportunity to carry at least some of the aggregate mortality risk themselves. Tontines are available in the United States from the Colleges Retirement Equity Fund. See www.tiaa-cref.org/prospectuses/cref_prospectus/index.html#A024.

do not offer any view on the scale of this pressure[2]. Hardy (2005) states that "Even actuaries recognize that longer life is a good thing – but, to the extent that it is unanticipated, it is also an enormous problem for the managers of annuity portfolios", with the implication that an annuitant should expect to pay a substantial risk premium on an annuity. On the other hand studies such as Finkelstein and Poterba (2002) argue that, after allowing for adverse selection, annuity pricing is close to actuarially fair before making any allowance for a risk premium. This last observation implies that either the risk premium cannot be very large or that, once it is taken into account, annuities are substantially mis-priced.

One clue to the pricing of mortality risk might be offered by financial markets. In November 2004 the European Investment Bank (EIB) issued a mortality bond following an earlier issue in December 2003 by the insurance company Swiss Re. A mortality bond is a loan stock whose payout depends on the mortality rate of a specific cohort and in the case of the EIB issue, the mortality of the cohort of men in the United Kingdom aged 65 in 2003. The pay-out received by purchasers of the stock increases proportionately to the longevity of the defined cohort, thus providing a hedge against aggregate mortality risk. Friedberg and Webb (2006) explore the pricing of the EIB mortality bond in the context of the consumption capital asset pricing model (CCAPM). Despite their observation that annuities expose insurance companies to "substantial risk", they find on the basis of their model that the cost of hedging aggregate mortality risk should be very low.

The CCAPM prices assets taking into account correlations between associated investment returns and consumption needs. This is an appealing approach for pricing annuities as it reflects the insurance motive. The CCAPM does not, however, provide an explanation of pricing in terms of demand and supply. Furthermore, the assumption of an infinitely lived representative consumer by the CCAPM sits uncomfortably alongside the uncertain mortality that influences demand for annuities in practice.[3]

Annuities represent a transaction, where one group of consumers (the elderly) divest themselves of their aggregate mortality risk by purchasing insurance from another group (the young). In this paper we consequently explore the pricing of aggregate mortality risk in an overlapping-generations model. In the model young people have no risk of death but, beyond a threshold age when they become old, they face a constant risk of death. Old people buy annuities from young people (or

[2]In the United Kingdom there is a separate problem that mortality rates have been systematically over-predicted by the Government Actuary. Here, however, we focus on shocks relative to unbaised forecasts.

[3]The model could be extended to address aggregate mortality risk using the approach suggested by Yaari (1965) in which all consumers have the same mortality rate independent of their age and where, therefore, the notion of a representative consumer can be retained, as Blanchard (1985) does in his analysis of fiscal policy with finite horizons.

from insurance companies whose shares are owned by young people) at a price which balances the willingness of the young to carry aggregate mortality risk with the desire of the old to divest themselves of it. The old and the young have different attitudes to the mortality risk of the old because they are of different ages; young people can adjust their consumption while still young in response to the gains or losses that they experience from carrying the mortality risk of the old. The fact that different generations are affected differently by shocks to the mortality rate of any particular cohort makes transaction in mortality risk possible.

We begin by discussing the way in which life expectancy has changed in the last twenty-five years or so. We follow this with estimates of the uncertainty surrounding the life expectancy of a sixty-five year-old man. We proceed to set out a modelling framework to assess the impact of aggregate mortality risk on the pricing of annuities and we use this to assess the risk margins which we might expect to observe. Finally we draw conclusions.

2 Rising Life Expectancy

The nature of the problem faced by insurance companies selling annuities can be seen from the data in figure 1. This shows official UK estimates of the cohort life expectancy for both men and women aged sixty-five in each year from 1981 to 2004. That life expectancies for both sexes have risen since 1981 is not surprise. The rise in male life expectancy came after a period of stasis lasting from about 1950 to 1975 which Doll (1987) attributes to the timing of the cigarette epidemic and this may help explain why the rise in male life expectancy has since been steeper than that of female life expectancy.

Rising but predictable life expectancy would not be a problem for insurance companies selling annuities. The difficulty arises because cohort life expectancy depends on future mortality rates and thus cannot be predicted with certainty. Indeed all the data shown in figure 1 depend on an element of prediction. Some of the people who reached age sixty-five in 1981 are still alive; the total expected life of people born in 1916 is obviously not yet known with certainty. There is very obviously much greater uncertainty about the expected remaining life of people born in 1939. Nevertheless these official figures are not presented with any indication of the uncertainty surrounding them. The problem faced when selling annuities is that such figures cannot be taken at their face value. The expected life of someone currently age sixty-five depends on forecasts of mortality rates for the next forty-five years or so and the terms on which annuities are traded might be expected to reflect the accuracy of these forecasts.

Given the time lags involved in the calculation of cohort life expectancy at age sixty-five, it is

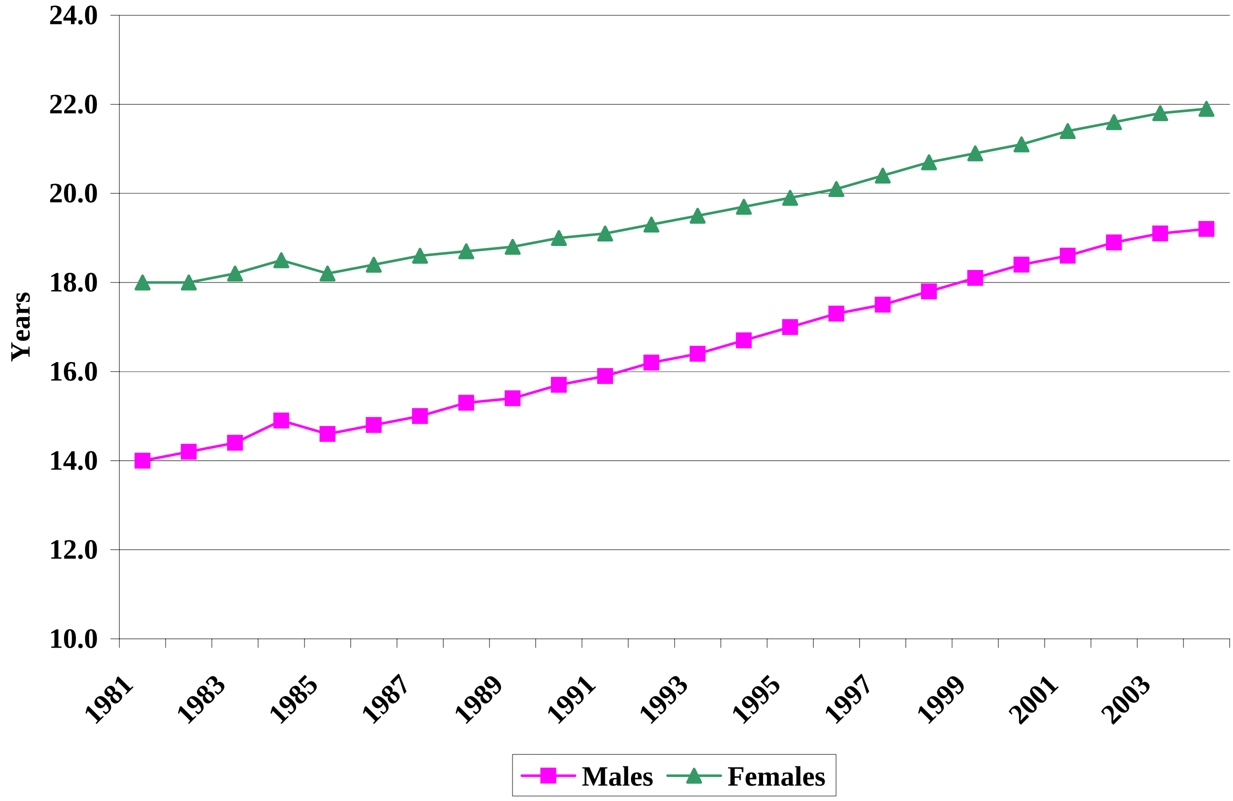

Figure 1: Estimates of Cohort Life-Expectancy at Age Sixty-Five: 1981-2004

not really practical to base any notion of the reliability of forecasts (i.e. those estimates produced before the cohort has died completely) on past experience of the magnitudes of forecast errors. Equally, plausible models of disturbances to life expectancy are not analytically tractable. Thus the uncertainty surrounding current estimates of life expectancy has to be established by means of stochastic simulation. We do this using the Lee-Carter (Lee and Carter 1992) model[4].

3 Stochastic Analysis of Life Expectancy

The Lee-Carter Model provides a simple framework for projecting future mortality rates. The basic model underlying mortality is assumed to be

$$\log m(x,t) = a(x) + b(x)k(t) + \varepsilon(x,t)$$

[4]See also Renshaw and Haberman (2003).

where $m(x,t)$ is the mortality rate for someone of age x in year t. $a(x), b(x)$ and $k(t)$ are model parameters and $\varepsilon(x,t)$ are error terms. The parameters are estimated to minimise $\sum_{x,t} \varepsilon^2(x,t)$ subject to the constraints that $\sum_t k(t) = 0$ and $\sum_x b(x) = 1$. Forecasts of mortality rates are produced by forecasts of k outside the sample and Lee and Carter show that a stochastic analysis can be carried out effectively by focussing on the uncertainty surrounding the projections of k; we follow that approach here.

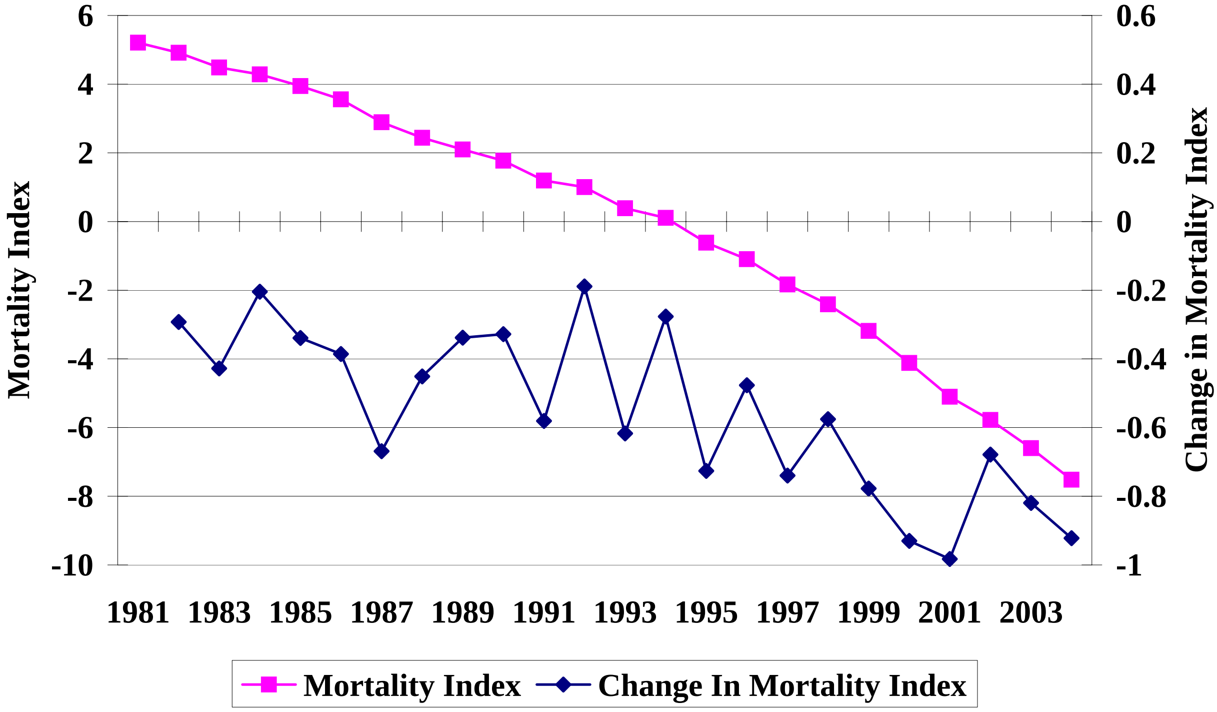

Figure 2: The Level and Difference of the Mortality Index for Men aged Sixty-five and Over:: 1981-2004

The index and its first difference are shown in figure 2 calculated from the publicly-available mortality rates for men which go back to 1981. The series itself is strongly trended and even its first difference does not appear stationary. The mean rate of change for 1982-1991 is -0.40 with the standard deviation of the mean 0.046. For 1995-2004 the mean is -0.76 with standard deviation 0.056. Thus the rate of reduction of the mortality index was substantially greater in the last ten years of the data than the first ten years and by any reasonable measure the difference is statistically significant. An assumption that the first difference followed a stable process would be

Variable	Order	ADF
k	0	3.63
Δk	1	-1.34
$\Delta^2 k$	2	-4.35

Table 1: Unit Root Tests for the Mortality Index

likely to deliver poor forecasts. If we look at ADF tests with the lag length chosen by the AIC criterion, we find the results shown in table 1. k needs to be double-differences in order to find a stationary process.

Although in principle one could estimate a time-series model for k or Δk and rely on the estimation method to deliver the restrictions which would be equivalent to estimating a model defined over $\Delta^2 k$ in practice the short sample makes this a highly risky procedure. It is much safer to consider only models set out in terms of $\Delta^2 k$. We explore autoregressive equations with up to three lags and choose the equation with three lags on the basis of the AIC. This equation is shown in table 2. The initial equation, although it passes a full range of specification tests, embodies an unstable negative root; the sum of the three coefficients on the lagged terms is well below -1. We therefore consider a restricted form, also shown in the table, in which the coefficients are restricted to sum to -0.9. We also restrict the constant term to zero. These restrictions are accepted by an F-test, F(2.14)=1.84.

	Unrestricted			Coeffs sum to -0.9		
	Coefficient	Std.Error	t-value	Coefficient	Std Error	t-value
$\Delta^2 k(-1)$	-0.77582	0.2331	-3.33	-0.5462	0.1706	-3.31
$\Delta^2 k(-2)$	-0.46127	0.3127	-1.47	-0.0456	0.0874	-0.52
$\Delta^2 k(-3)$	-0.54981	0.2492	-2.21	-0.2902	0.1706	-1.72
Constant	-0.0808	0.04472	-1.81	0	0	**
σ	0.16985			0.1902		
R^2	0.659269					
F(3,14)	9.029					
log-likelihood	8.63204					

Table 2: An Autoregression for $\Delta^2 k$

This gives us one model from which to produce both forecasts of cohort life expectancy and, using stochastic simulation, estimates of its dispersion. When we simulate stochastically the autoregressive model, there is a risk that some of the sets of stochastically chosen parameters for the autoregression will result in models which are dynamically unstable. In 1000 simulations we find that 142 result in unstable models and have to be discarded. From the remaining simulations we find a mean value for life expectancy of those currently aged sixty-five of 19.8 years with a standard deviation of 0.73 years. This compares with an official estimate of 19.4 years and a high variant of 20.3 years.

However, at the same time we have to note that these results are, even though the simulations take account of the uncertainty in the model parameters, conditional on the specification of the model being correct. Clements and Hendry (1999, Chapter 5) suggest that an alternative robust forecasting model is often provided by the assumption that $\Delta^2 k$ is a random variable with zero expected value. The negative terms in the regression equation suggest that this alternative specification is likely to have greater uncertainty associated with it than does the model based on the regression equation. This model gives, again from 1000 simulations, a male life expectancy at age sixty-five of 20.0 years with a standard deviation of 1.67 years. With this background we now look at a formal structure for modelling the supply and demand for annuities, so as to work out the market-clearing annuity rate.

4 The Decision Framework

Our model structure focuses on two overlapping generations which we refer to as young and old. Young people are begin their economic life at age 0 and all of them survive to age 1 when they become old. One period can be thought of as representing forty years. All old people in each cohort experience a cohort-specific mortality rate which is independent of age beyond the age of 1, but which is not known with certainty until after they have to make financial decisions for the period starting at age1. The mortality rate in beyond age 1 for cohort i is defined as ρ_i. The analysis of annuities presented by Yaari (1965) applies to the old generation with the qualification that at the age of 1 they can buy annuities before their mortality rate is known. Alternatively they can buy tontines whose pay-out depends on their actual mortality rate once it is established. The interest rate in the model is r and the rate of discount of future utility is also assumed to be r.

Cohort i faces three decision points in our model.

1. At age 0 young people in cohort, i, receive an endowment of £1. They can sell annuities to old people from cohort $i-1$ who have now reached the age of 1. However, they will choose to do so only at a price which reflects the uncertainty surrounding the mortality of cohort $i-1$; the expected mortality rate of cohort $i-1$ is ρ^e_{i-1}. We denote by π_{i-1} the ratio of the actual annuity pay-out to that which would be made by a fair annuity without uncertainty. Thus an annuity bought by an old person in cohort $i-1$ at age 1 pays a dividend of $\pi_{i-1}\left(r+\rho^e_{i-1}\right)$. In

the absence of uncertainty a fair annuity bought for £1 would yield a dividend of $r+\rho_{i-1}$. The capital value once the uncertainty is revealed is consequently $£\frac{\pi_{i-1}\left(r+\rho^e_{i-1}\right)}{r+\rho_i}$ per £1 invested. The expected capital gain is therefore[5]

$$£1 - \pi_{i-1}(r+\rho^e_{i-1})E\left(\frac{1}{r+\rho_{i-1}}\right).$$

Given the pay-out ratio in the market and the uncertainty surrounding the mortality rate of old people in cohort $i-1$, young people in cohort i decide to sell annuities to a value of $£\phi'_i$.

2. At a time ε after this initial transaction, the actual mortality rate, ρ_{i-1}, of the old cohort becomes known. Given the realisation of ρ_{i-1} and thus the capital gains or losses that young people from cohort i have made, they choose their consumption rate c^y_i between the ages ε and 1. The assumption that the discount rate equals the interest rate means that this consumption rate is constant throughout the interval $[\varepsilon, 1]$.

3. At age 1 cohort i becomes old. It now holds wealth w^1_i and has to decide what proportion of this, ϕ_i it wishes to annuitise. It makes this decision in the light of the uncertainty surrounding its own mortality rate, ρ_i, and in the face of the market pay-out ratio for annuities, π_i. It buys a tontine with its remaining wealth[6] giving a yield based on the as yet unknown mortality rate. This determines the income and consumption of the cohort when old, c^o_i.

We assume that ε is small so that we can ignore the consumption which takes place over the interval $[0, \varepsilon]$.

With an interest rate of r,

$$w^1_i = \left\{1+\phi'_i\left(1-\pi_{i-1}\frac{r+\rho^e_{i-1}}{r+\rho_{i-1}}\right)\right\}e^r - \frac{c^y_i}{r}\left(e^r-1\right) \tag{1}$$

An old person with wealth w^1_i at time 1 who invests a proportion ϕ_i in an annuity and the remainder in a tontine receives a level income of $y^o_i = \{(1-\phi_i)(r+\rho_i) + \phi_i\pi_i\left(r+\rho^e_i\right)\}\, w^1_i$. Within the confines of our model, the results of Yaari (1965) and Davidoff, Brown, and Diamond (2005) imply that old people will invest all of their wealth in one or other of these products; they also imply that, provided the interest rate equals the discount rate, which we assume, these investors

[5] As the arithmetic mean is greater than the harmonic mean for any sequence of positive variables, it follows that for zero expected capital gain, $\pi_{i-1} < 1$. In other words, an annuity which is fair in the presence of uncertainty would seem unfair if assessed without taking that uncertainty into account.

[6] With our assumptions buying a tontine at time 1 is equivalent to buying an annuity at time $1+\varepsilon$.

will choose consumption, $c_i^o = y_i^o$. Putting these expressions together allows us to write the budget constraint:

$$\begin{aligned} c_i^o &= \left\{ (1-\phi_i)(r+\rho_i) + \pi_i \phi_i (r+\rho_i^e) \right\} w_i^o & (2) \\ &= \left\{ (1-\phi_i)(r+\rho_i) + \pi_i \phi_i (r+\rho_i^e) \right\} & (3) \\ & \quad \left\{ \left(1 + \phi_i' \left[1 - \pi_{i-1} \frac{r+\rho_{i-1}^e}{r+\rho_{i-1}} \right] \right) e^r - \frac{c_i^y}{r} (e^r - 1) \right\} \end{aligned}$$

A situation in which tontines are not available is easily represented by setting[7] $\phi_i = 1$, with the consequence that people choose first ϕ_i' and then c_i^y

Each consumer has an instantaneous utility function

$$u(c_i^z) = A + \frac{(c_i^z)^{1-\alpha}}{1-\alpha}; \quad z = y \text{ or } o$$

Here A is a variable which represents the joy of living and is large enough to ensure that $u(c_i^z) > 0$ for all plausible values of c_i^z. This condition has to be met for consumers to regard long life as a good thing rather than a burden but does not influence utility-maximising decisions..

5 Optimal Allocation

Looking at the four decision points we work backwards, in common with the backward recursion used for solving dynamic programming problems. We define $V_{i,t}$ as the expected utility of someone in cohort i at time t on the assumption that they make optimal portfolio and consumption decisions. We set out the analysis on the assumption that tontines are available.

1. After the true mortality rate for old people is revealed at time $1+\varepsilon$, there is no further uncertainty and
$$V_{i,1+\varepsilon} = \frac{u(c_i^o)}{r+\rho_i}$$
2. At time 1 the consumer has to choose how much wealth to annuitise in order to maximise expected utility. It is assumed that old people cannot be sellers of either tontines or annuities;

[7] If the alternative to an annuity is cash with a yield of r, then there is the logical possibility that people would wish to allocate their portfolios between annuities and cash. However, exploration of this possibility showed that, for any plausible mortality rate, people would choose to borrow at the market interest rate in order to increase their annuity holdings. Such an analysis assumes that estates are simply insolvent on death. In such circumstances lenders would require borrowers to buy life insurance and the attractions of over-annuitisation would be lost.

$0 \leq \phi_i \leq 1$.

$$
\begin{aligned}
V_{i,1}\left(\pi_i, w_i^1, \rho_{i-1}\right) &= \begin{array}{c} Max \\ 0 < \phi_i < 1 \end{array} E\left(V_{i,1+\varepsilon}\right) \\
&= \begin{array}{c} Max \\ 0 < \phi_i < 1 \end{array} \frac{(w_i^1)^{1-\alpha}}{1-\alpha} \\
& \quad \int_0^\infty \frac{\left\{(1-\phi_i)(r+\rho_i) + \pi_i\phi_i\left(r+\rho_i^e\right)\right\}^{1-\alpha}}{r+\rho_i} f\left(\rho_i, \rho_{i-1}, \Delta\rho_{i-1}\right) d\rho_i \\
& \quad + \int_0^\infty \frac{A}{r+\rho_i} f\left(\rho_i, \rho_{i-1}, , \Delta\rho_{i-1}\right) d\rho_i \qquad (4)
\end{aligned}
$$

We represent the density function of ρ_i as depending on the values of ρ_{i-1} and $\Delta\rho_{i-1}$ to take account of the persistence of mortality rates (See section 6). We stress that this is a function of the annuity pay-out, π_i

3. At time ε the rate of consumption when young must be chosen. However no new information accrues in the interval $[\varepsilon\ 1]$ and therefore

$$
V_{i,\varepsilon}\left(\pi_i, w_i^\varepsilon, \rho_{i-1}, \Delta\rho_{i-1}\right) = \frac{A\left(1-e^{-r}\right)}{r} + \begin{array}{c} Max \\ c_i^y \end{array} \left\{\frac{\left(1-e^{-r}\right)}{r\left(1-\alpha\right)}(c_i^y)^{1-\alpha} + e^{-r}V_{i,1}\left(\pi_i, w_i^1, \rho_{i-1}, \Delta\rho_{i-1}\right)\right\} \qquad (5)
$$

4. At time 0 the consumer must decide how far it is prepared to sell annuities to the current elderly. π_i is unknown because it reflects the demand for annuities. The homotheticity of our utility functions makes clear that this depends on the amount of wealth that consumers own when they reach age 1, w_i^1. This in turn depends on the profits that they make from selling annuities when aged 0 and thus on the realisation of ρ_{i-1}. ϕ_i' is found by solving

$$
V_{i,0} = \begin{array}{c} Max \\ \phi_i' \end{array} E\{V_{i,\varepsilon}\left(\pi_i, w_i^\varepsilon, \rho_{i-1}, \Delta\rho_{i-1}\right)\}
$$

We set $w_i^\varepsilon = 1 + \phi_i'\left(1 - \pi_{i-1}\frac{r+\rho_{i-1}^e}{r+\rho_{i-1}}\right)$. The homothetic form of

$$
V_{i,\varepsilon}\left(\pi_i, 1, \rho_{i-1}, \Delta\rho_{i-1}\right) - A\left\{\frac{1-e^{-r}}{r} + \int_0^\infty \frac{f\left(\rho_i, \rho_{i-1}, \Delta\rho_{i-1}\right)}{r+\rho_i} d\rho_i\right\}
$$

means that, with

$$
U_i\left(\rho_{i-1}, \Delta\rho_{i-1}\right) = A\left\{\frac{1-e^{-r}}{r} + \int_0^\infty \frac{f\left(\rho_i, \rho_{i-1}, \Delta\rho_{i-1}\right)}{r+\rho_i} d\rho_i\right\}
$$

$$
\begin{aligned}
V_{i,0} &= \begin{matrix} Max \\ \phi_i' \end{matrix} \; E\left\{\left(1+\phi_i'\left[1-\pi_{i-1}\frac{r+\rho_{i-1}^e}{r+\rho_{i-1}}\right]\right)^{1-\alpha}\left(V_{i,\varepsilon}\left(\pi_i,1,\rho_{i-1},\Delta\rho_{i-1}\right)-U_i\left(\rho_{i-1},\Delta\rho_{i-1}\right)\right)\right\} \\
&\quad +U_i \\
&= \begin{matrix} Max \\ \phi_i' \end{matrix} \int_{\Delta\rho_{i-1}}\int_{\rho_{i-1}}\int_{\pi_i}\left(1+\phi_i'\left[1-\pi_{i-1}\frac{r+\rho_{i-1}^e}{r+\rho_{i-1}}\right]\right)^{1-\alpha} \\
&\quad (V_{i,\varepsilon}[\pi_i,1,\rho_{i-1},\Delta\rho_{i-1}]-U_i)h(\pi_i)g(\rho_{i-1},\Delta\rho_{i-1})d\pi_i d\rho_{i-1}d\Delta\rho_{i-1} \\
&\quad +U_i
\end{aligned}
$$

where $g\left(\rho_{i-1},\Delta\rho_{i-1}\right)$ is the joint density function of ρ_{i-1} and $\Delta\rho_{i-1}$ conditional on the state of knowledge when cohort i is aged 0 and $h\left(\pi_i\right)$ is the density function of π_i conditional on the same information. Since the realisation of ρ_{i-1} determines π_i through the market-clearing process we can write

$$
\begin{aligned}
V_{i,0} &= \begin{matrix} Max \\ \phi_i' \end{matrix} \int_{\rho_{i-1}}\left(1+\phi_i'\left[1-\pi_{i-1}\frac{r+\rho_{i-1}^e}{r+\rho_{i-1}}\right]\right)^{1-\alpha} \\
&\quad \left\{V_{i,\varepsilon}\left(\pi_i[\rho_{i-1},\Delta\rho_{i-1}],1,\rho_{i-1},\Delta\rho_{i-1}\right)-U\left(\rho_{i-1},\Delta\rho_{i-1}\right)\right\}f\left(\rho_{i-1},\Delta\rho_{i-1}\right)d\rho_{i-1}d\Delta\rho_{i-1} \\
&\quad +\int_{\rho_{i-1}}U\left(\rho_{i-1},\Delta\rho_{i-1}\right)d\rho_{i-1}d\Delta\rho_{i-1}
\end{aligned}
$$

If tontines are not available then the analysis proceeds as set out above, but with the restriction that $\phi_i = 1$.

In either case, an obvious difficulty is that we cannot expect to find any functional form for $\pi_i(\rho_{i-1},\Delta\rho_{i-1})$

6 Mortality Shocks

In common with empirical work on the topic, we assume that the mortality rate is log-normally distributed with density function $f\left(\rho_i\right)$. Thus

$$\log\rho_i = \frac{\gamma_1+\nu_i}{\delta}$$

where ν_i is a normally distributed error term. Following the analysis of section 3 we assume that

$$\Delta\nu_i = \Delta\nu_{i-1}+\upsilon_i \quad \text{where } \upsilon_i\tilde{}N(-\gamma_2,1), \quad Cov(\upsilon_i,\upsilon_j)=0 \;\; i\neq j$$

Working from ρ_{i-2} which is known with certainty at the time cohort i makes its first economic decisions, this gives

$$\begin{aligned}\log \rho_{i-1} &= \log \rho_{i-2} + \frac{\nu_{i-2} + \upsilon_{i-1}}{\delta} \\ \log \rho_i &= \log \rho_{i-2} + \frac{2\nu_{i-2} + 2\upsilon_{i-1} + \upsilon_i}{\delta}\end{aligned}$$

Setting $\gamma_2 = 1/2\delta$ ensures that, if $\nu_{i=2} = 0$, $E\left(\rho_{i-1} \mid \rho_{i-2}\right) = \rho_{i-2}$. Nevertheless, the asymmetry in the log-normal distribution means that $E\left(\rho_i \mid \rho_{i-2}\right) > \rho_{i-2}$ even though $E\left(\rho_i \mid \rho_{i-1}\right) = \rho_{i-1}$.We carry out our analysis with $\nu_{i-2} = 0$.

We assume that the first time period, is equivalent to forty years and that the expected mortality rate in the second period is 5% p.a. giving a mortality rate of 2 per forty-year period. Thus

$$E\left\{e^{\frac{\gamma_1+\nu_i}{\delta}}\right\} = 2 \tag{6}$$

and the life expectancy on reaching retirement is $E\left\{e^{-\frac{\gamma_1+\nu_i}{\delta}}\right\}$ which is slightly above 1/2. Equation (6) provides one restriction needed to establish γ_1 and δ. We also assume that

$$Var\left\{e^{\frac{\gamma_1+\nu_{i-1}}{\delta}}\right\} = 0.04$$

in order to determine γ_1 and δ uniquely with $\gamma_1 = 6.8989$ and $\delta = 10.0249$. In other words the standard deviation of the mortality rate is 1/10 of its expected level. This then yields a value of ? for γ_2.The standard deviation of life expectancy at age sixty-five is on the high side compared to the analysis of section 3, but some degree of mark-up is reasonable since the analysis there does not take account of factors such as model uncertainty. The assumption of log-normality implies that we can use the Gaussian quadrature to compute expectations when we calculate the optimal choices. These parameter values give us a life expectancy of 0.5050 or 20.2 years in retirement.

7 The Market Solution

We set out here demand and supply curves for annuities. We look first at the situation where tontines are available.

7.1 The Demand for Annuities

The demand curve for annuities is given as the value of ϕ_i which optimises equation (4). The homotheticity of the utility function implies that this is independent of w_i^o. Given our specification of ρ_i the integral can be evaluated using Gaussian quadrature; we use five abscissae with the assumption that $\nu_{i-1} = 0$. The optimum is then found using *fminsearch* in MATLAB. However, we note a tendency to over-annuitisation. With $\pi_i = 1$ we show in appendix A working to a second-order approximation, that $\phi_i = \frac{\alpha+1}{\alpha}$. Since this is independent of the magnitude of $Var(\rho_i)$ provided the latter is positive it is due to some fact other than the consequences of the differences between arithmetic and harmonic means noted above.

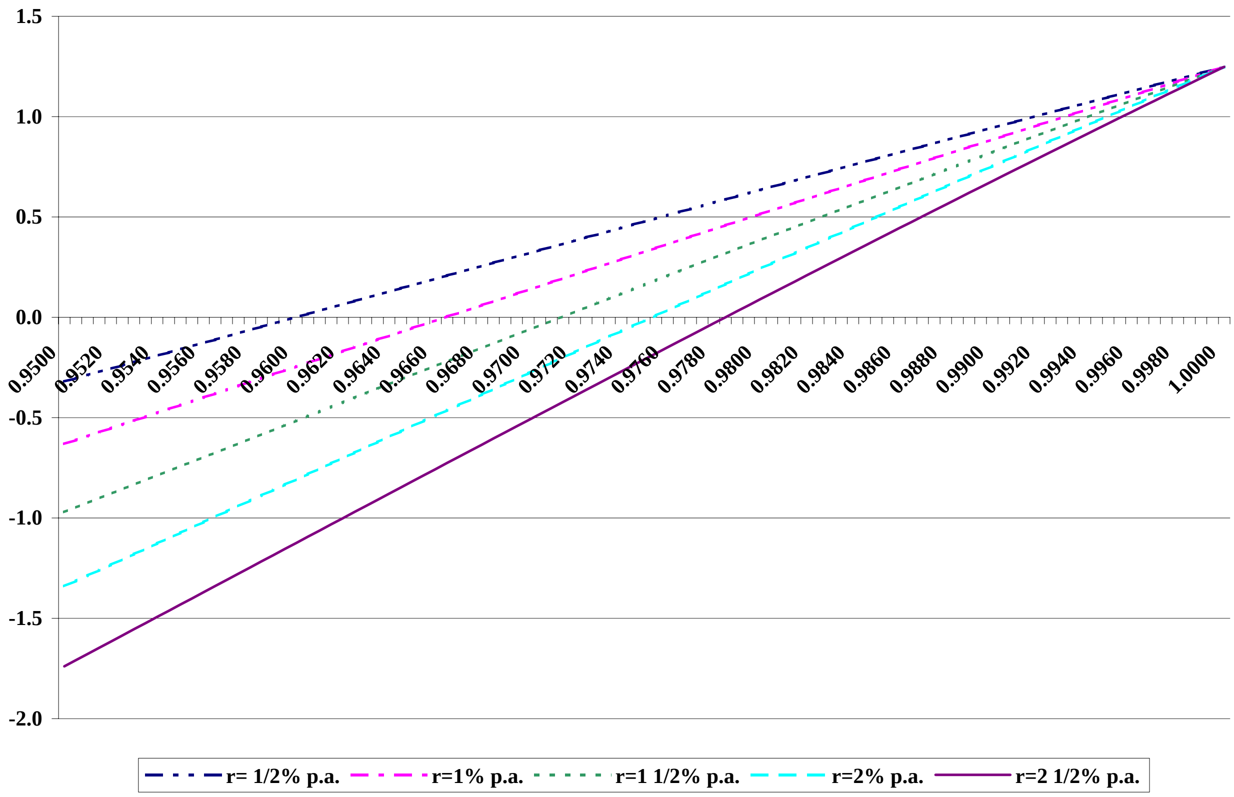

Figure 3: Demand for Annuities: Relative Risk Aversion = 4

The proportionate uncertainty in consumption arising from uncertainty in ρ_i is decreasing as a function of r. If the interest rate is low then the overall return on a tontine, or on funds annuitised at age $1+\varepsilon$ is dominated by ρ_i; it follows that shocks to ρ_i will have a large proportionate impact and people will be keen to buy annuities even if the pay-out is poor. If r is large, then both ρ_i and shocks to ρ_i are likely to be less important. As a consequence, at any given value of $\pi_i < 1$ the proportion of w_i^1 which old people wish to annuitise will be a declining function of r. We illustrate demand curves for five values of r between $\frac{1}{2}\%$ p.a. and $2\frac{1}{2}\%$ p.a. in figure 1. The calculations are carried out on the assumption that an age of 1 represents forty years, so that r is forty times its annual rate. We use a value of $\alpha = 4$. This is towards the upper range of plausible estimates for this value; we choose a value which implies a high degree of risk aversion because the aim of this paper is, within the framework of a structural model, to look at the effects of risk aversion on the supply of and demand for annuities.

The curves confirm that the steepness of the demand curve is a function of the interest rate, for the reasons we discussed above. At low annuity rates it is not surprising that old people would like to be net sellers rather than net buyers of annuities. But the chart suggests that, for an annuity market to exist when investors can choose between annuities and tontines, the value of π_i cannot be very far below 1. If sellers of annuities need to make a substantial charge for risk, then there should not be an annuity market, and its existence can be explained only as a result of the legal requirement for pension funds to be used to purchase annuities rather than tontines.

If tontines are not available, all wealth is annuitised and the demand curve is therefore vertical.

7.2 The Supply of Annuities

While the demand curve for annuities can be drawn unambiguously, the supply curve depends on the uncertainty that is associated with future pay-out rates. The homotheticity of the utility function implies that wealth w_i^1 when reaching age 1 is an increasing function of the profit realised on the sale of annuities when young. Since the proportion of wealth annuitised at any pay-out ratio is independent of the amount of wealth, a high level of wealth will raise total demand for annuities and thus depress the pay-out ratio. Thus the response of the pay-out ratio dampens, at least to some extent, the effect of uncertain returns from the sale of annuities, compared to a situation where the pay-out ratio is unresponsive. To clarify the discussion here, we assume that the annuity pay-out consumers expect when old will be the same as that used to derive the relationship between current pay-out and supply; in other words $g(\pi_i) = \delta(\pi_i - \pi_{i-1})$ where $\delta(x)$ is the delta function with $\delta(x) = 0$ if $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x) = 1$. A family of these supply curves is shown as a function of the interest rate. It should, however, be borne in mind that, when the effects of the uncertainty in the subsequent pay-out ratio are taken into account, the proportion of initial wealth engaged in the sale of annuities will be higher than is shown in figure 4.

The supply curve for annuities is also, of course, sensitive to the persistence of mortality shocks across cohorts. In order to demonstrate the effects of this we present two families of supply curves. The first is constructed with the assumption that mortality shocks are serially independent while in the second we assume that the mortality shocks follow the second-order process described in section 6.

We find again that the steepness of the curve is increasing in the interest rate. This once again reflects the fact that, with high interest rates mortality risk is quantitatively less important than

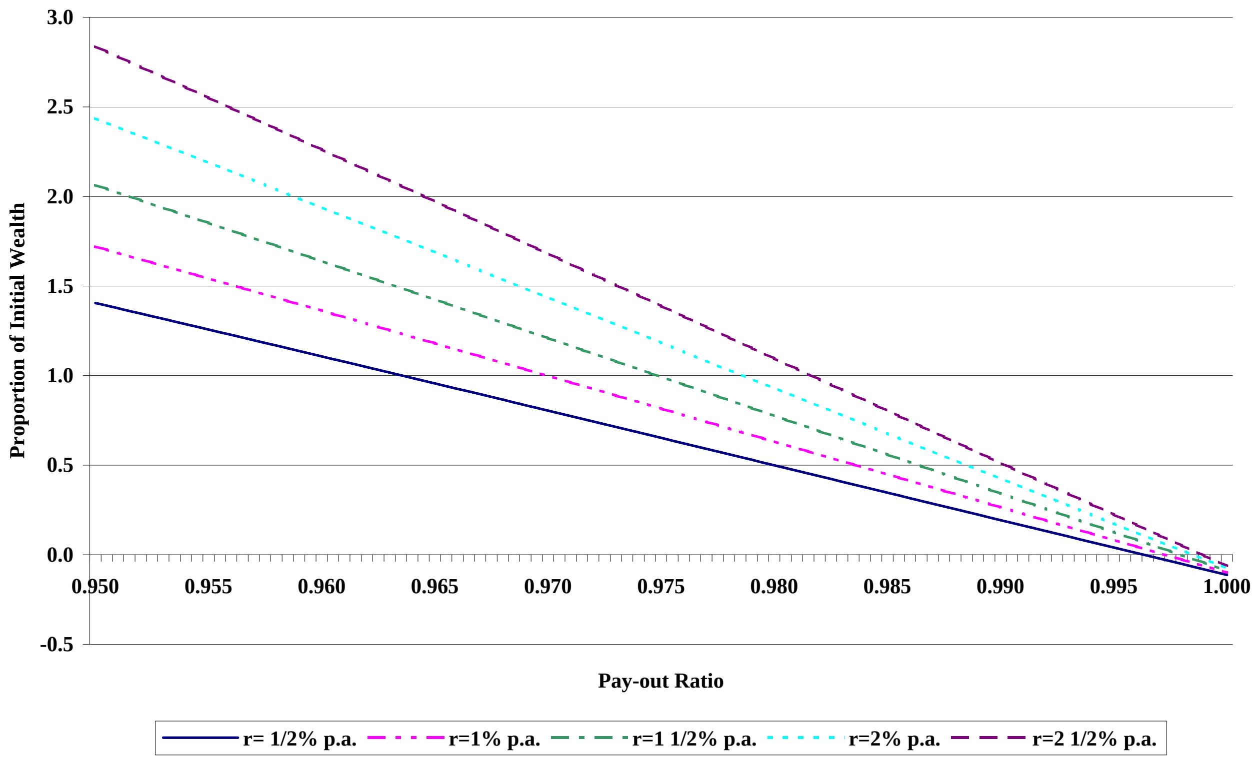

Figure 4: Supply of Annuities : Relative Risk Aversion = 4, No persistence of mortality shocks

with low interest rates. In consequence the proportion of wealth people are prepared to annuitise rises more steeply as the pay-out on the annuity falls. With a pay-out of 1 young people want to be buyers rather than sellers of annuities (selling a negative proportion of initial wealth). This is a consequence of the fact that, with uncertainty, the fair pay-out ratio is below 1. Once again the higher is the interest rate, the less important is mortality uncertainty and thus the fair pay-out ratio is closer to 1.

When we look at the effects of persistence of mortality shocks we see the rather different picture shown in figure 5. When young investors lose money because their parents live for longer than expected, the persistence of mortality rates mean that, in addition they have a longer expected period of retirement to support. Thus the financial effect of losses arising on the sale of annuities is magnified compared with the situation where mortality rates are independent. Thus it is not very surprising that the supply curves are both shallower and shifted to the left as compared to the position where mortality rates are independent. The effect is at its most marked when the interest rate is very low for the reason identified earlier. When interest rates are very low the capitalised effects of shocks to mortality are greater than when they are higher.

When the mortality rate of each cohort is independent of those of the other cohorts the supply curve of annuities is not affected by whether retired people have to annuitise all of their wealth or whether they can by tontines. This property, confirmed by simulation, follows from the homothetic nature of the utility function. But when mortality rates are serially correlated the separation property which leads to this result no longer holds. We show in figure 6 the supply curves when tontines are not available. As compared to figure 5 these curves have moved to the left but are steeper, so that at high pay-out ratios the supply of annuities is reduced while at low pay-out ratios it is increased.

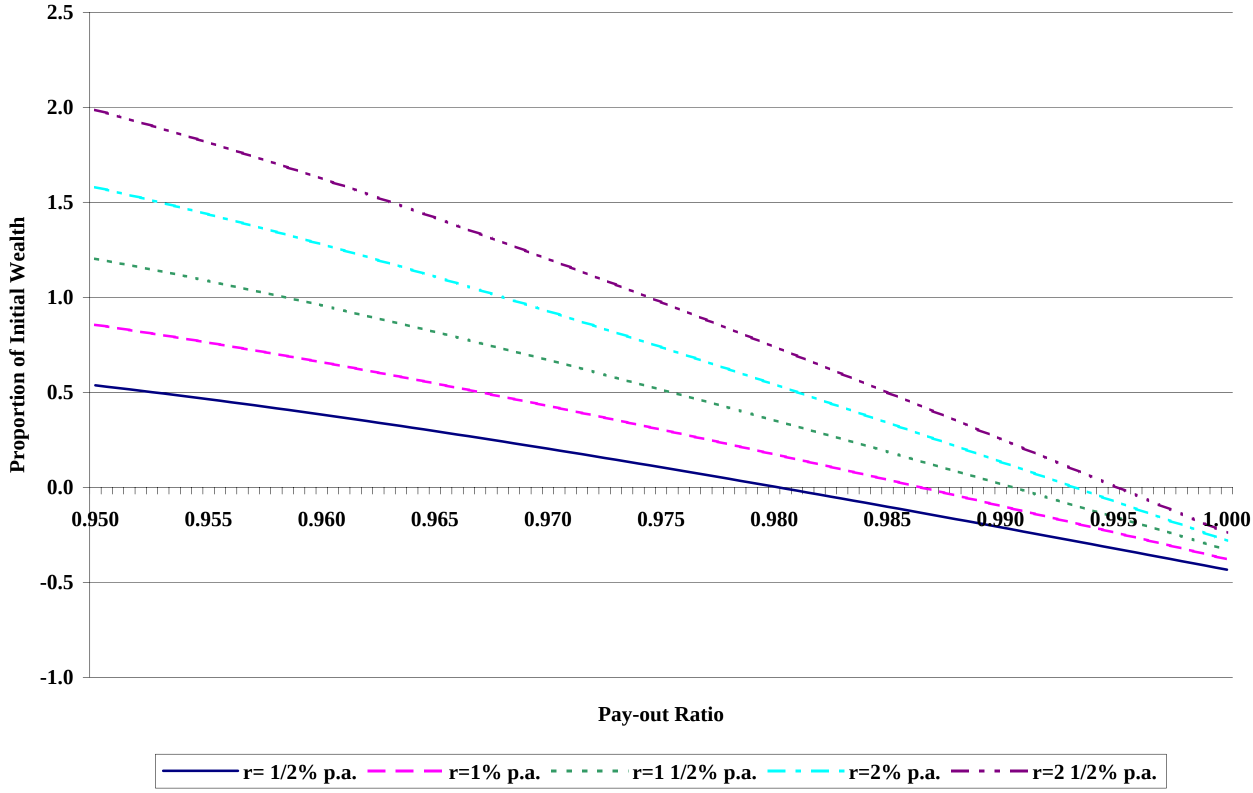

Figure 5: Supply of Annuities : Relative Risk Aversion = 4, Mortality Shocks follow 2nd Order Process, Tontines available

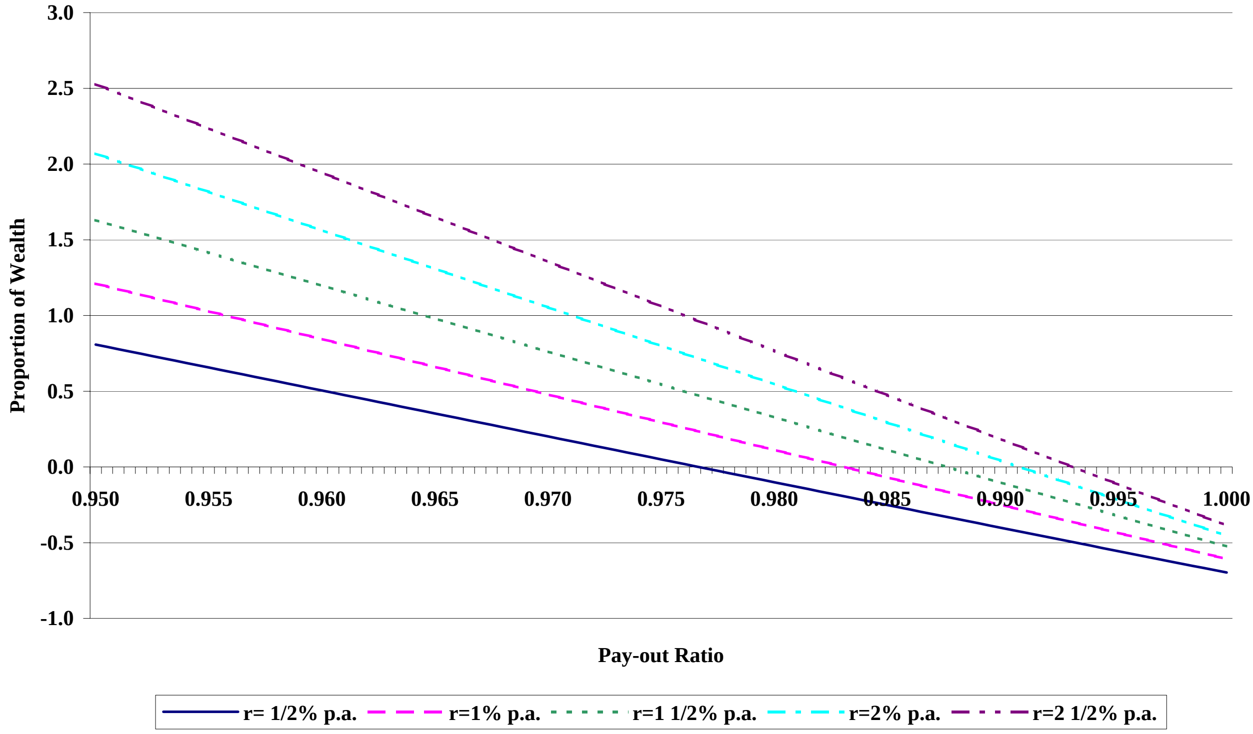

Figure 6: Supply of Annuities : Relative Risk Aversion = 4, Mortality Shocks follow 2nd Order Process, Tontines not available

8 Market Equilibrium

8.1 Solution Method

As with any structural model of demand and supply, we can identify the market-clearing price- in this case we focus on its reciprocal, the pay-out ratio. We simulate a market-clearing process with the price adjusting so that the share of old people's wealth which young people are prepared to annuitise is equal to the proportion that old people would like annuitised. The reciprocal of the pay-out ratio can be thought of as the risk premium associated with demographic uncertainty.

The equilibrium condition is

$$\phi_i' \left(\pi_{i-1}, \pi_i \left(\nu_{i-1}\right), g[\pi_i(\pi_{i-1}, \Delta\pi_{i-1})]\right) = \phi_{i-1} \left(\pi_{i-1}\right) w_{i-1}^1(\nu_{i-2})$$

with the supply of annuities shown to be a function of the current mark-down, π_{i-1}, the pay-out in the next period which is a function of ν_{i-1}, the shock to the mortality rate of the current elderly, and the density function of the pay-out in the next period, $g[\pi_i(\pi_{i-1}, \Delta\pi_{i-1})]$ which may itself depend on the current pay-out. The wealth of the current elderly is shown as a function of the mortality shock which affected the cohort immediately preceding them, ν_{i-2}. If $g[\pi_i(\pi_{i-1}, \Delta\pi_{i-1})]$ and $\pi_i \left(\nu_{i-1}\right)$ were known, one would use a simple tatonnement to find, π_{i-1}. Using the superscript j to denote the jth iteration

$$\pi_{i-1}^{j+1} = \pi_{i-1}^{j} + \lambda \left\{\phi_i' \left(\pi_{i-1}^{j}, \pi_i \left(\nu_{i-1}\right), g[\pi_i(\pi_{i-1}, \Delta\pi_{i-1})]\right) - \phi_{i-1} \left(\pi_{i-1}^{j}\right) w_{i-1}^{o}(\nu_{i-2})\right\}$$

ensures that when supply exceeds demand the pay-out is raised, making annuities more attractive to buyers and less attractive to sellers. Experimentation is needed to find a value of λ which ensures rapid convergence.

The wealth that old people bring to the market is a function of the profits that they make on trading in annuities when they are young as described by equation (1). We therefore develop a first estimate of the density function of π_i, $g^0 \left(\pi_i, \Delta\pi_{i-1}\right)$, by finding the values of $\pi_{i-1} \left(\rho_{i-1}, \Delta\rho_{i-1}\right)$ generated for different realisations of ν_{i-1}, the disturbances to $\log\rho_{i-1}$ on the assumption that π_i takes the value π_{i-1} and that this is known with certainty. In other words

$$g^1 \left(\pi_i \left(\rho_{i-1}, \Delta\rho_{i-1}\right)\right) = \delta \left(\pi_i - \pi_{i-1}^1 \left(\rho_{i-1}, \Delta\rho_{i-1}\right)\right)$$

where $\delta \left(x\right)$ is once again Dirac's delta-function.

We now use $g^1 \left(\pi_i \left(\rho_{i-1}, \Delta\rho_{i-1}\right)\right)$ to iterate on each of the values of ν_{i-1} required for the Gaussian quadrature calculations as

$$\pi_{i-1}^{2,j} \left(\rho_{i-1}, \Delta\rho_{i-1}\right) = \lambda \left\{\phi_i' \left(\pi_{i-1}^{2,j}, g^1[\pi_i^1 \left(\rho_{i-1}, \Delta\rho_{i-1}\right)]\right) - \phi_{i-1} \left(\pi_{i-1}^1\right) w_{i-1}^{o} \left(\rho_{i-1}, \Delta\rho_{i-1}\right)\right\}$$

and use the resulting converged values, $\pi_{i-1}^{2,n}\left(\rho_{i-1}\right)$ to define the grid representing $\pi_i^2\left(\rho_{i-1}\right)$. We iterate, evaluating

$$\pi_{i-1}^{k,j}\left(\rho_{i-1},\Delta\rho_{i-1}\right)=\lambda\left\{\phi_i'\left(\pi_{i-1}^{k,j},g^{k-1}[\pi_i^{k-1}\left(\rho_{i-1},\Delta\rho_{i-1}\right)]\right)-\phi_{i-1}\left(\pi_{i-1}^{k-1}\right)w_{i-1}^o\left(\rho_{i-1},\Delta\rho_{i-1}\right)\right\}$$

until the converged values, $\pi_{i-1}^{k,n}\left(\rho_{i-1}\right)$ are close enough to $\pi_{i-1}^{k-1,n}\left(\rho_{i-1}\right)$ for all values of ν_{i-1} . At this point we have identified, for each value of ν_{i-1} a value of π_{i-1} which clears the annuity market, on the assumption that the density function $g\left(\pi_i\right)$ is consistent with the pattern of disturbances to ρ_{i-1}.

8.2 Results

The structure we have set out allows us to compare the market equilibria and their implications in four different cases. We look at the situation when tontines can be bought and also when they are not available, and we also compare the implications of a stationary and serially independent mortality rate with one where the mortality rate is stable in second differences. We present in table 3 for the situation where cohort i is selling annuities to cohort $i-1$, showing the market-clearing mark-down on the pay-out in the absence of uncertainty, π_{i-1}, the propensities to invest in tontines when young and when old (ϕ_i' and ϕ_{i-1}), the expected propensity to consume out of wealth when young, c_i^y, evaluated at time 0 for cohort i, the expected life-time utility of the cohort (omitting the contribution made by the term A) and relative to the case where tontines are available and where the mortality rate is stationary and serially uncorrelated for which it is defined as 1, the level of initial wealth which is needed to deliver the same level of utility as in that case. Thus this last variable, denoted Y_i allows us to identify the cost of one situation relative to another. However these results do not fully take account of the effects of the uncertainty that young sellers of annuities face about the future mark-down on annuity rates when they become buyers on reaching retirement and for this reason they are subject to revision.

This table draws attention to a number of interesting points. First of all, in none of the solutions is the mark-down more than 3% on the annuity which would be paid in the absence of mortality risk. Thus, if the market is efficient, we should not expect to see mortality risk having a large effect on annuity rates. When mortality rates are stationary and not serially correlated, the absence of tontines has the effect of depressing the pay-out by 0.3% with small implications for consumption and welfare. When mortality follows the second-order process the impact on the market-clearing rate is larger, reflecting the movement in the supply curves shown above.

Head	Tontines Available	Tontines not Available
Mortality Stationary		
π_{i-1}	0.9897	0.9868
ϕ_i'	0.2763	0.3837
ϕ_{i-1}	0.7233	1
c_i^y	0.9035	0.9028
Mortality Second Order		
π_{i-1}	0.9796	0.9724
ϕ_i'	0.1324	0.3890
ϕ_{i-1}	0.3455	1
c_i^y	0.9025	0.9003

Table 3: Market Equilibrium

9 Conclusions

A general equilibrium framework allows us to explore the demand for annuities, on the assumption that people can choose between these and tontines as a means of providing for old age. With plausible assumptions about mortality risk, the rate of return and the degree of risk aversion, the pay-out ratio of an annuity which compensates annuity sellers for the risks that they are carrying is over 97% of the pay-out of a fair annuity, giving a risk premium of less than 3%. even in the most adverse of the circumstances we have investigated. The analysis shows clearly that there is a demand for tontines because retired people would not rationally wish to be fully protected from the risks associated with aggregate mortality uncertainty. Thus a policy goal should be to stimulate the development of tontines as products available to old people as a means of financing retirement.

References

Banks, J., and R. Blundell (2005): "Private Pension Arrangements and Retirement in Britain," *Fiscal Studies*, 26, 35–54.

Blanchard, O. (1985): "Debt, Deficits and Finite Horizons," *Journal of Political Economy*, 93, 223–247.

Bohn, H. (2005): "Who bears what Risk? An Intergenertional Perspective," www.econ.ucsb.edu/ bohn.

Clements, M., and D. Hendry (1999): *Forecasting Non-stationary Economic Time Series*. The MIT Press, Cambridge, Massachussets.

Davidoff, T., J. Brown, and P. Diamond (2005): "Annuities and Individual Welfare," *American Economic Review*, 95, 1573–1590.

Doll, R. (1987): "Major Epidemics of the 20th Century: from Coronary Thrombosis to AIDS," *Journal of the Royal Statistical Society: Series A*, 150, 373–395.

Finkelstein, A., and J. M. Poterba (2002): "Selection effects in the United Kingdom individual annuities market," *Economic Journal*, 112, 28–50.

Friedberg, L., and A. Webb (2006): "Life is Cheap: Using Mortality Bonds to Hedge Aggregate Mortality Risk," NBER Working Paper No 11984. www.nber.org/papers/W11984.

Hardy, M. (2005): "A Matter of Life and Death," Financial Engineering News: www.fenews.com/fen46/topics act analysis/topcis-act-analysis.htm.

Lee, R., and L. Carter (1992): "Modelling and Forecasting US Mortality," *Journal of the American Statistical Association*, 87, 659–671.

Renshaw, A., and S. Haberman (2003): "Lee-Carter Mortality Forecasting: a Parallell Generalized Linear Modelling Approach for England and Wales Mortality Projections," *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 52, 119–137.

Yaari, M. (1965): "Uncertain Life-time, Life Insurance and the Theory of the Consumer," *Review of Economic Studies*, 32, 137–150.

A Over-annuitisation

It is simple to demonstrate that, in the absence of a risk premium old people tend to over-annuitise. We set $r + \rho_i = R_i$ and $r + \rho_i^e = R_i^e$. Then

$$
\begin{aligned}
c_i^o &= \{(1-\phi_i) R + \pi_i \phi_i R^*\} w_i \\
\frac{\partial c_i^o}{\partial \rho_i} &= (1-\phi_i) w_i \\
\frac{\partial c_i^o}{\partial \phi_i} &= \pi_i R^* - R
\end{aligned}
$$

and

$$E(u_i^o) = u(c_i^o) + \frac{\sigma^2}{2}\left\{\frac{2u(c_i^o)}{R^3} - \frac{2u'(c_i^o)}{R^2}(1-\phi_i) w_i + \frac{u''(c_i^o)}{R}(1-\phi_i)^2 w_i^2\right\}$$

Maximising this with respect to ϕ_i

$$\begin{aligned}\frac{\partial E\left(u_i^o\right)}{\partial \phi_i} &= \left(u\left(c_i^o\right)+\frac{\sigma^2}{2}\left\{\frac{2u\left(c_i^o\right)}{R^3}-\frac{2u'\left(c_i^o\right)}{R^2}\left(1-\phi_i\right)w_i+\frac{u''\left(c_i^o\right)}{R}\left(1-\phi_i\right)^2 w_i^2\right\}\right)\left(\pi_i R^*-R\right)\\ &\quad +\frac{\sigma^2 u'\left(c_i^o\right)}{R^2}w_i-\frac{\sigma^2 u''\left(c_i^o\right)}{R}w_i^2\left(1-\phi_i\right)\end{aligned}$$

If $\pi_i = 1$ and $R^* = R$ then the first order condition is

$$\sigma^2 u'\left(c_i^o\right) = \sigma^2 u''\left(c_i^o\right) R w_i \left(1-\phi_i\right) \tag{7}$$

With the utility function given by (**??**) this suggests the solution

$$\phi_i = \frac{\alpha+1}{\alpha} > 1 \text{ with } \alpha > 0$$

which is independent of the mortality risk. The result appears to suggest that overannutization takes place even in the absence of aggregate mortality risk. However if $\sigma^2 = 0$, then it is of course not permissible to divide both sides of (7) by σ^2. It does, however indicate that with values of π_i close to but below 1 there is likely to be over-annuitisation.