A Conceptual Framework For Retirement Products: Risk Sharing Arrangements Between Providers And Retirees

Gregorio Impavido, Craig Thorburn, Mike Wadsworth†

The World Bank and Watson Wyatt

Abstract

Voluntary annuity markets are in most countries smaller than what the theoretical and part of the empirical literature would suggest. There are both demand and supply constraints that hamper the development of annuity markets. In particular, traditional products available in most countries can require excessive minimum capital requirements for given investment opportunities available to providers. Investment and longevity risk should be shared between providers and annuitants so that supply constraints can be relaxed. Alternative annuity products, which imply risk sharing, could be backed by substantially lower capital investments or, equivalently, provided at substantially lower prices to consumers.


The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at http://econ.worldbank.org.

JEL Classification Codes: G22, G23.

Keywords: Annuities, Retirement benefits, Risk sharing.

† Gregorio Impavido (gimpavido@worldbank.org) is Senior Financial Economist and Craig W. Thorburn (cthorburn@worldbank.org) is Senior Financial Specialist, both in the Financial Sector Operations and Policy Department of the World Bank. Mike Wadsworth (mike.wadsworth@eu.watsonwyatt.com) is Partner at Watson Wyatt. The authors are grateful to Mike Orszag and Sara Zervos for useful comments.
INTRODUCTION

The low level of development of annuity markets around the world is of particular concern in the light of increasing number of pension reform proposals that promote prefunding of pension liabilities in the private sector and the increasing appeal of defined contribution schemes for sponsors. Traditional annuity products seem not to be popular among consumers and in many countries retirement benefits based on self-insurance of longevity risk are common. The content of this paper should naturally appeal to policy makers concerned with old-age poverty and the adequacy of retirement benefits. It should appeal to an academic audience interested in the development of private annuity markets, saving decisions and wealth allocation by households. It should also appeal to reformers of pension systems encouraging savings accumulation and financing of annuities through private sector arrangements. It is the opinion of the authors that, within the trend of prefunding and private sector solutions, insufficient consideration is given to the ability of private sector arrangements to supply adequate annuities to retirees.

The objective of Section I of this paper is to identify demand constraints for annuity products that policy-makers can attempt to relax with appropriate institutional and regulatory design. In particular, we present the predictions of Yaari’s (1965) model, report on the size of country-specific annuity markets, summarize the empirical literature on the value of annuities for individual consumers, and conclude with considerations on whether the low demand for voluntary annuities can indeed be considered “puzzling.” In Section II we focus on possible supply constraints on the part of annuity providers. The objective of this section is to posit that another possible cause for the low size of voluntary annuity markets is that annuity providers bear higher than desirable levels of risk for the given products offered and the availability of assets that can be used to match liabilities.1 We define and discuss several types of risk that providers typically bear when selling traditional, or usually available, annuity products. We present a conceptual framework for sharing these risks between providers and annuitants. We suggest that alternative annuity products, which imply risk sharing, could be backed by substantially lower capital investments or, equivalently, provided at substantially lower prices to consumers. Conclusions follow in Section III.

I ANNUITY DEMAND CONSTRAINTS: THE ANNUITY PUZZLE

This section provides a survey of the existing economic literature on household annuity decisions.2 In particular, we focus on the contrasting predictions of the theoretical literature and available empirical evidence on voluntary asset annuitization by households during retirement: i.e., on the so called “annuity puzzle.”3

---

1 The extent of such mismatch is the object of empirical investigation and a natural candidate for future research.
2 The literature on annuities, as it is strictly connected to the literature on savings and consumption behavior during retirement age, is very rich and a comprehensive literature survey on the subject falls beyond the scope of this paper.
3 This section extends the presentation framework that can be found in Brown (2001).
I.A Yaari’s (1965) model

The seminal contribution of Yaari (1965) is often referred to as the first paper to consider in a life-cycle savings model the effect of random timing for the terminal condition on individuals’ decision to purchase annuities. The model developed assumes that: 1) consumers are Von Neumann-Morgenstern expected utility maximizers; 2) consumers’ preferences are time independent; 3) the only risk faced by consumers is longevity risk; 4) complete insurance for this type of risk is available through annuities; 5) annuities are actuarially fair\(^4\) and that therefore, they pay a rate of return higher than the market rate on conventional assets;\(^5\) 6) there is only one conventional asset which pays a given interest rate; and 7) consumers can borrow and lend at this same rate.

Yaari (1965) considers four different states of the world depending on whether consumers have access to insurance markets or not and whether consumers have bequest motives or not. The results are reported in the following table.

Table 1: Yaari’s (1965) results

<table>
<thead>
<tr>
<th>Access to Insurance</th>
<th>Bequest Motive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>(A) When consumers have no access to longevity insurance they fully bear longevity risk. Their subjective discount rate at time (t) can be decomposed in two summands: a) the subjective discount rate that would prevail at time (t) with no survival uncertainty; and b) the probability of living at time (t). Consumption is depressed in a world where longevity risk exists compared to a world with no longevity risk. Also, due to the assumption of no bequest motive, consumers are constrained to have non-negative assets with probability one at the time of death. (i.e.,) bequests, if any, are unintentional.</td>
</tr>
<tr>
<td>Yes</td>
<td>(C) As far as the rate of consumption is concerned, the introduction of insurance at fair value is equivalent to removing longevity risk from the problem. Since by assumptions annuities (actuarial notes) yield an interest rate that is higher than other interest bearing assets, consumers holding assets would do so only in the form of annuities. In this way, negative net worth can never be attained during the life-cycle and bequests are zero with probability one at death.</td>
</tr>
</tbody>
</table>

Case C and D are the relevant ones. In a life-cycle framework and given assumptions 1) to 7) above, consumers will be better off by holding only annuity assets if they have no bequest motives, while, when they have a bequest motive, they will hold a portfolio of

---

\(^4\) Yaari (1965) defines “an actuarial note as a note which the consumer can either buy or sell and which stays on the books until the consumer dies, at which time it is automatically cancelled” (pag. 140). The purchase of an actuarial note coincides with the purchase of an annuity whilst the sale of an actuarial note coincides with the purchase of a life insurance policy (more precisely, a life insured loan with, due to the actuarial fairness assumption, repayment rate higher than the market rate until death and no further obligation).

\(^5\) If an annuity is actuarially fair it would pay, gross of administrative costs, a premium on conventional assets due to the presence of different longevity types in the pool.
annuities and bequeathable assets so that the marginal utility of bequests and consumption are the same.

In the next section we report on the size of annuity markets in specific countries to provide factual evidence on whether consumers indeed behave according to the predictions of the life-cycle hypothesis as presented here.

I.B Size of annuity markets in different countries

A consistent cross-country analysis of the size and growth of annuity markets is not yet available. This is mainly due to the non-negligible variance of pension arrangements and to the very small size of voluntary individual annuity markets around the world. Cardinale et al. (2002) is a good summary of cross country institutional arrangements for retirement income provision but it does not always provide the quantitative comparison on annuity markets that is relevant for this section: in particular, the role of life annuities within households’ net worth. Despite these shortcomings, it is fair to infer from the survey in this section that, despite their recent increased importance and their long existence,6 annuities (and furthermore individual voluntary life annuities) neither do they represent the totality of retirees’ savings, nor any considerable portion.

For the United States, Poterba (2001) reports various indicators of growth of annuities: 1) annuity payouts, as a percentage of life insurance payouts, increased from 7% to 40% in the period between 1940 and 1999; 2) premiums for individual annuity policies and group annuities increased in real terms by more than 85 and 32 times, respectively, during the period between 1951 – 1999; 3) individual annuity premiums increased from 0.064% of GDP in 1951 to 1.2% of GDP in 1999; 4) group annuity premiums grew from 0.23% of GDP in 1951 to 1.6% of GDP in 1999; and 5) total annuity reserves increased from less than 50% of total life insurance reserves in 1960s to more than twice the value of total insurance reserves in the 1990s. The rapid growth of annuities in the States is attributed to concerns about financial stability, increased per capita disposable income, and the growth of corporate pension plans. However, despite its growth, the annuity market in the United States and in other countries remains low.

Mitchell and Moore (1988) report for the US that around 60% of total household wealth is constituted by private pensions and social security pensions while the other 40% represented by bequeathable assets. The market of annuities, as reported by the authors, is not negligible. As a matter of fact, in 1998, premiums paid for single premium immediate individual annuities were US$ 7.9 billion, premiums for immediate group annuities were US$ 16.3 billion and premiums for deferred annuities were US$ 117.7 billion. However, the authors quote that a large proportion of these contracts do not have longevity insurance attached. Also, that the vast majority of individual annuities are purchased as a consequence of settlements legal cases and therefore, do not represent savings for retirement. Finally, their definition of annuity market includes deferred annuities, from which it is often possible to withdraw assets without conversion to a life-long income stream. Hence, Brown et al. (2002) conclude that the market for

---

6 James (1947) reports that even during Roman times there existed contracts, called annua, that exchanged a stream of income for a fixed or variable period of time with an up-front payment.
retirement-linked annuities with lifetime insurance\(^7\) is still very small, with annual premiums in the neighborhood of US$ 2 billion in 1998. Despite the current small size of the individual annuity market, Brown \textit{et al.} (2002) believe that the market is likely to grow in the future due to the growth of retirees participating in defined contribution accounts. Mitchell \textit{et al.} (1999) also argue that there is a strong growth potential for annuity markets and that considerable growth has already taken place in the variable annuities market. However, most of these annuities are still in the accumulation phase and, like deferred annuities, regulation allows savers not to annuitize. Hence, it is not possible to safely infer from variable annuities growth patterns just reported that growth is related to retirement needs.

Finkelstein and Poterba (2002), Murthi \textit{et al.} (1999) and Brown \textit{et al.} (2001) describe the institutional structure of the annuity market in the UK. The market for pensions annuities is larger than the market for non-pension annuities\(^8\) and a large series of products is sold in both markets.\(^9\) Within the pension annuity market, the compulsory market is very large with annuity payments totaling £3.9 billion in 1997. Annual payments to voluntary annuitants in 1996 were only £0.8 billion. Premiums for immediate individual annuities amounted to £4.2 billion in 1996 while Cardinale \textit{et al.} (2002) report that the total premiums for pensions and voluntary annuities amounted to around £8 billion in 2001. In other words, the UK immediate annuity market is much larger than the US$ 2 billion US individual immediate annuity market as reported by Brown \textit{et al.} (2002).\(^{10}\)

Bateman and Piggott (1998, 1999 and 2002) provide an excellent discussion of retirement income provision in Australia.\(^{11}\) The Old Age pension, by definition, insures against longevity risk and in 2001, around 80% of retirees received some form of Old Age pension. This is set as a minimum of 25% of male average earnings and represents

\(^7\) Namely, individual, immediate, single premiums, life annuities.

\(^8\) The UK annuity market is divided into pension and non-pension annuities. Pension ('compulsory purchase') annuities may be written by the pension provider or bought on the 'open market' from another provider. Non-pension annuities consist of annuities typically purchased voluntarily with assets accumulated through general savings.

\(^9\) The menu is very rich: level annuities pay a constant sum of money; impaired life and/or enhanced annuities link the base payment to the characteristics of the annuitant; unit-linked annuities transfer the investment risk to the annuitant and guarantee payments defined as a number of units, rather than in a absolute monetary value; index-linked annuities are basically unit linked annuities but are “linked” to some price index (e.g., the retail price index); with-profits annuities pay an amount linked to the profits of the provider; flexible annuities allow annuitants do define how and when to draw income and allow control on investment portfolio. All these can be single life or joint; with or without minimum guarantee; with constant or variable annuity frequency.

\(^{10}\) And more concentrated too, with top 4 annuity companies representing 63 percent of the market. Cardinale \textit{et al.} (2002) argue that the concentration of the annuity market in the UK is due to the fact that: 1) many but few companies have drastically reduced their presence in the market after having written a lot of business based on overoptimistic assumptions; 2) economies of scale in the provision of annuities; and 3) there are significant competitiveness differentials between the internal and external business.

\(^{11}\) The Australian pension system can be divided in three “pillars” A first pillar represented by a means tested (cash) social safety net financed from the budget. A second pillar of private managed mandatory savings represented by the Superannuation Guarantee. A third pillar of voluntary and/or tax preferred savings. Notice that since the first pillar is means tested, it also represents minimum state guarantee for the second pillar.
the only meaningful source of longevity insurance in Australia. The Superannuation Guarantee does not provide insurance against longevity risk as there is no regulation on the form with which income is to be paid so that around 85% of benefits are paid in the form of a lump sum. James and Vittas (2000) argue that the growth of the annuity market in Australia is mostly due to the introduction of the Superannuation Guarantee. The market for allocated annuities\textsuperscript{12} is also growing rapidly, in partly due to the investment choice attached to these products. Knox (2000) reports that allocated annuities, term annuities, and life annuities represent around 60%, 28% and 12% of the annuity market, respectively. Total reserves for these products accounted for 15% of total life insurance reserves in 1998, up from 4% of in 1994. Cardinale \textit{et al}. (2002) report that new premium income for individual life annuities was around US$ 94 million in 2001. As in all countries so far surveyed, the traditional life annuities market is very small in Australia.

In India, James and Sane (2002) report that the annuity market is small but has been growing quickly. The annuity premiums growth in recent years has been uneven but it has grown from 0.065% to 2.9% of total life insurance premiums between 1996 and 2000. In 2000, the number of annuity policies were still only 1.3% of life business in terms of new policies. Individual annuity business is a large group business from superannuation plans, which include pension investments and annuities. Annuities are generally viewed in the country as tax-advantaged saving measures instead of means to ensure old-age security.

Palacios and Rofman (2001) review the annuity market experience from four Latin American countries: Argentina, Chile, Colombia and Peru. Retirement benefits from the defined contribution component of each pension system can take the form in all four countries of phased withdrawals and annuities.\textsuperscript{13} The annuity markets in Argentina, Peru and Colombia are very small with gross premium income less than 0.2% of GDP. The Chilean market is relatively large, with annuity gross premium income in the neighborhood of 1.5% of GDP. One of the main reason attributed by the authors to the difference in market size is the earlier date of the Chilean pension reform. Premium written by specialized annuity companies in Argentina was around US$ 156 million, 18% of total life insurance premium income, in 1998. In Chile, annuity premiums were around US$ 1.7 billion in 1997, figure comparable in absolute terms to the retirement-linked individual annuity market of the United States as reported by Brown \textit{et al}. (2002). Annuity premiums in Colombia amounted to around US$ 18 million in 1998.

In other countries statistics are harder to obtain. In Germany, Schnabel (2002) reports that private savings account for around 10% of disposable income but that German household portfolios are concentrated on housing wealth and other bequeathable assets such as bank accounts, life insurance, stocks and bonds, while private annuities are

\textsuperscript{12} Strictly speaking, these are not life annuities but phased withdrawal schemes with allowed maximum that sees to guarantee income until age 80 and allowed minimum based on life expectancy. In other words, an individual withdrawing the minimum allowed is self-insuring the longevity risk that in average, although only on average, he is not outliving his own resources.

\textsuperscript{13} For instance, regulation in Chile requires retirees either to receive benefits from AFPs in the form of phased withdrawals, or in the form of an immediate annuity from a specialized annuity company, or as a combination of a deferred annuity with phased withdrawals (James and Vittas 2000).
negligible. Cardinale et al. (2002) report that annuities represent only 5% of benefits paid by life insurance companies as of 1999 but that the share of annuity premiums increased from 5.1% to 22.6% of total life insurance products between 1991 and 1999. In Italy, Cardinale (2002) and Cardinale et al. (2002) also report that the annuity market is very small. Net wealth of Italian households is estimated at around twice Italian GDP but like German households, Italian household portfolios are concentrated on housing. Other financial assets are distributed among bank deposits (57%), bonds (14%), equities (6%), life insurance (4%) and other financial instruments (17%). Premiums for annuity products have actually decrease in the period 1996 – 1998 from 10% to 6% of total life insurance products. In Singapore, Doyle et al. (2001) argue that the introduction in the 1990s of a requirement for a minimum sum to be held in individual accounts in the Central Provident Fund has contributed to the development of the annuity market as evidenced by an increase from 380 to 3200 annuities sold per year from 1990 to 1999. However, the market remains small as only one sixth of the retired population purchased annuities for a total premium of Sg$ 173 million (US$ 105 million) in 1999.

The country specific information reported in this section seems to suggest either of the two following conclusions, whether consumers’ have bequest motives or not:

1. In the absence of bequest motives: a) the predictions of the Yaari’s (1965) model, that lifecycle consumers should annuitize all their wealth, are not supported by casual observation. Hence, either the life cycle hypothesis is incorrect, or assumptions made in the literature are too stringent.15

2. In the presence of bequest motives: a) the utility of bequest is sufficiently high; and b) consumers are constrained in trading across states of the world between annuitized and traditional assets so that marginal utilities of these two assets cannot be equated. Insurance markets incompleteness is the likely constraint here. Hence, if both 2a) and 2b) hold at the same time, voluntary annuity markets remain very small.16

I.C The value of annuities

The question of why annuities markets are so small has generated a large empirical literature on the value of annuities for individual consumers. The book by Brown et al. (2001) contains the core papers, by the same authors, on three major annuity valuation indices: the money’s worth ratio (MWR), the wealth equivalence (WE), and the annuity equivalent wealth (AEW). The MWR is a non utility based measure that relates the expected present value of future stream of income from an annuity to the present value of the sum its policy premiums. Ratios below (above) 1 would indicate that consumers value future income streams from annuities less (more) than the premiums paid. The WE is a utility based indicator that measures the amount of numeraire that non-annuitized consumer would need to use to achieve the same indifference curve in a fully annuitized

---

14 Chia and Tsui (2003) argue that there is a significant spread in the value of retirement consumption between males and females and that the minimum sum should be raised by some 2% to meet female retirees consumption needs.
15 In Section I.D we review possible relaxations of the Yaari (1965) assumptions.
16 Altruism and insurance market incompleteness are considered in Section I.D as two of the possible explanations for why the predictions of the life cycle model appear not to be observed in reality.
The AEW is a utility based measure the amount of numeraire that a fully annuitized individual would need to achieve the same indifference curve in a non-annuitized world.\(^{18}\)

The first measure is easiest to implement as it does not involve any assumption on the shape of individuals’ indifference curves. All three measures require assumptions on: 1) the individual discount rate; 2) survival probabilities; and 3) annuity payouts. The complete terms structure of interest rate should be used to derive individuals’ discount rates as these vary with time and age. Obviously, unsatisfactory assumptions need to be made in less developed financial markets where bonds with sufficiently long duration are usually not issued. Cohort mortality tables for annuitants should be used to derive survival probabilities in order to endogenize longevity improvements\(^{19}\) and eliminate the effect of adverse selection.\(^{20}\) Less developed countries hardly have population mortality tables and level less frequently have cohort or easily available annuitant tables. As in the case of individual discount rates, other unsatisfactory assumptions need to be made when using these measures in less developed financial markets.

Mitchell et al. (1999) use the MWR and the WE frameworks to evaluate non-participating, single-premium, immediate, nominal,\(^{21}\) individual life annuities in the US.\(^{22}\) In their calculations, the authors: 1) use the term structure of Treasury and corporate\(^{23}\) bonds interest rates to discount future annuity payouts; 2) consider the effect of tax treatment of annuities to distinguish between before-tax and after-tax interest rates; and 3) use both cohort mortality tables for the population as a whole and for the annuitant population. They find that: 1) there exists a substantial variation in the annuity benefits paid by different insurers; 2) the expected present value of annuity payouts to a man aged 65 in 1995 was valued between 81% and 93% of the annuity premium,\(^{24}\) implying that consumers buy a considerably less than fair insurance product; 3) the internal rates of return of annuities are between 1% and 2% below that yielded by Treasury or corporate bonds, mirroring the result in 2); and 4) that the negative margin on annuity premiums

---

17 For a risk averse consumer and a given level of utility, the insurance value of an annuity would cause, ceteris paribus, WE to be lower than the non annuitized wealth.
18 Essentially, WE and AEW are similar to the money metric indirect utility functions that are constructed by means of the expenditure function in welfare analysis. Notice, that the EW and AEW would be one the reciprocal of the other in a world where individuals have a linear utility function and where, therefore, the ratio of the two would yield the marginal rate of substitution between annuitized and non-annuitized assets.
19 Period mortality tables report the mortality experience of a person at a given point in time. Cohort mortality tables report the mortality experience of a person at a given point in time, contingent to his/her year of birth (also called dynamic mortality tables). Cohort mortality tables incorporate longevity improvements of specific cohorts so that the survival probability at time \(t\) of an individual born at time \(t - n\) is generally lower than the survival probability at time \(t + n\) of an individual born at time \(t\).
20 See later on the discussion on adverse selection.
21 Or in British tradition “level”.
22 The argument made for focusing on this specific subset of annuities is that these are the only type of annuities that are unambiguously linked to the insurance of longevity risk during retirement.
23 Although constructed by adding a constant to the Treasury bonds yield curve.
24 The difference is due to the use of population or annuitant mortality tables, and of the Treasury or corporate yield curve. Ceteris paribus, expected present values of annuities are higher when annuitant mortality tables are used as annuitants have a higher life expectancy than the population as a whole. They are also higher when the Treasury yield curve is used, as this implies assuming a lower subjective discount rate.
has been decreasing during the last decade, suggesting improved efficiency of the market. The WE of annuities for a 65 years old man, with no pre-annuitized wealth and different assumptions on real interest rate, risk aversion and if utility is considered pre or after-tax, is between 60% and 70% of non annuitized wealth in a world with certain inflation. The same measure is between 70% and 80% when consumers have 50% of wealth pre-annuitized. The same figures are slightly higher in a world of uncertain inflation.

Brown et al. (2001) use the MWR and AEW measures to valuate nominal and real (retail-price indexed) single premium, immediate, life annuities in the UK, as well as real and variable annuities in the US. Differently from Mitchell et al. (1999), the authors use cohort mortality tables for the population as a whole and report always pre-tax valuations. They find that: 1) similarly to the Mitchell et al. (1999) for the US, there exists a substantial variation in the annuity benefits paid by different insurers; 2) for 65 years old male in the UK, nominal annuities with average payouts are valued at 91% of the annuity premium, while real annuities are valued at 85% of the annuity premium; 3) for a 65 years old male in the US, nominal annuities with average payouts are valued at 86% of the annuity premium, while ILONA real annuities are valued at 70% of the annuity premium, a good 15 percentage points lower than in the UK. The AEW for a 65 years old man in the US, with no pre-annuitized wealth and different assumptions on risk aversion, is between 150% and 200% of annuitized wealth, for real annuities, and 145% and 160% for nominal annuities. The same measures are between 133% and 181%, and 130% and 157%, respectively, when consumers have 50% of wealth pre-annuitized in real terms.

Finkelstein and Poterba (2002) evaluate the MWR of three types of monthly, single life annuity products in the UK: 1) nominal; 2) real; and 3) 5% escalating annuities. By using population mortality tables and annuities with average payouts, the authors report that for a 65 years old male, the average MWR is 90.1%, 82.2%, and 85.4% respectively for nominal, real and 5% escalating annuities in the compulsory annuity market. In the voluntary annuity market the same MWRs are 86.1%, 79.1% and 80.7%, respectively.

The results by Mitchell et al. (1999) indicate that despite a 65 years old man would expect to receive, in present value terms, between 81% and 93% of paid premiums for an annuity, his risk aversion make him value the insurance contract so much that he would be prepared to forego up to 40% of initial wealth to have access to the annuity market. The results by Brown et al. (2001) are complementary as they show that the same individual would need non-annuitized wealth starting in the neighborhood of 145% of his annuitized wealth to attain the same level of utility. Finally, the higher the level of risk aversion the lower the WE (the higher the AEW) generally needed to attain the same level of utility.

The review of the literature on the value of annuities, similarly to the theoretical literature based on the life cycle hypothesis, seems to indicate that consumers would be ready to pay a substantial premium to own these assets. Alternatively, that these assets

---

25 The real annuities valued in Brown et al. (2001) are a CPI-indexed annuity sold by the Irish Life Company of North America (ILONA).

26 Similarly, Poterba and Warshawsky (2001) report that in 1998 the average MWR for a 65 years old male was 84%, using population mortality tables.
should occupy a larger portion in household wealth than what is observed empirically. The next section attempts to reconcile these statements with country specific factual evidence and by considering possible relaxations of the assumptions made in Yaari (1965).

I.D Is low annuity demand really a puzzle?

The prediction of the lifecycle model that full annuitization is optimal in the absence of a bequest motive relies on a series of strong assumptions. One possible reaction to the so-called “puzzle” is to posit that the assumptions in Yaari (1965) are too strong for the prediction to be plausible. However, attempts to relax these assumptions have not provided a definite answer to the issue either. The literature summarized in this section provides very ambiguous answers on whether the annuity puzzle exists or not.

Various reasons not necessarily independent from one another have been identified for the low demand for annuities. Among these are low subjective life expectancy or myopia, adverse selection and high load factors, incomplete markets, pre-annuitized wealth, precautionary savings, and bequest motives.

The retirement risk survey conducted by the Society of Actuaries (SOA 2001) finds that men tend to underestimate average life expectancy at age 65, currently 81 to 83 years of age depending on the population projected. Women estimate their life expectancy only slightly better than men. Roughly half the women in the sample underestimate the average 65-year-old female’s life expectancy, which is 85 to 86 years of age. In general, only about one-third of retirees and pre-retirees are on target or err on the side of overestimating average life expectancy at age 65. Retirees and those nearing retirement age (age 45 and over) underestimate the life expectancy of the average 65 year old.

Adverse selection and loads are factors explaining money’s worth ratios below unity and are also considered to explain the annuity puzzle. The source of adverse selection lies in the difference between annuitant and population average mortality tables. Annuitants as a group tend to live longer than the population average. The presence of asymmetric information in insurance markets implies that individuals self-select themselves on the basis of private information about their longevity. This form of adverse selection can explain the high load factors in certain markets. For instance, Fong (2002) finds that in Singapore, adverse selection explains around 13% of the cost of longevity insurance and she concludes that a policy option would be to mandate annuitization to avoid the adverse selection problem. Murthi et al. (1999), in their price analysis of UK annuities conclude that the main source of reduction in annuity yields in that market is given by adverse selection that however, are not translated in high load factors. Finkelstein and Poterba (2002) find that private information not only affects the decision to participate in the insurance market, but also affects the choice of annuity products. For instance, annuitants who tend to live longer, also tend to select backloaded products. Finally, Warshwsky (1988) finds that an average load factor of 29 cents applies to the US annuity market but however, he concludes that this would not be sufficient to prevent individuals from buying annuities and that the low level of demand should be explained by other factors like bequest motives.
Incomplete insurance markets could account for the observed low demand for annuity products. When insurance markets are not complete individuals still bear some risk and the higher is their risk aversion the stronger is the need to use non-annuitized assets to insurge against such risks. For instance, nominal annuities do not insure against inflation risk and therefore, the monotonic relationship between risk aversion and AEW may not hold for high levels of risk aversion. A similar argument against the rationale for (full) annuitization is made by James and Vittas (2000) when considering the value of annuities for liquidity constraint consumers. The disutility from full annuitization is also discussed by Davidoff et al. (2003), although more formally. The extent with which this substitution effect takes place is, of course, left to empirical analysis.

The presence of pre-annuitized wealth obviously reduces the demand for annuities. Abel (1985) indicates that large public pensions effectively insure against longevity risk through intergenerational transfer of wealth. Minimum pension guarantees work in the same way. For instance, guarantees on phased withdrawals apply to Chile, Colombia, and Argentina. Chile guarantees also life annuity payments up to 100% of the minimum pension. In Argentina, annuities are guaranteed up to 1.6 times the average wage or five times the maximum basic pension. No explicit guarantee applies to retirement benefits in Peru (Palacios and Rofman 2001). However, Poterba (2001) does not conclude that, for the US, a reduction in annuitized benefits from social security would necessarily yield an increase in demand for private annuities.

As mentioned before, another important reason for the low demand for annuities is the bequest motive and a substantial part of the literature is dedicated to measuring the magnitude of the bequest motive. There are three main reasons for bequests: uncertain lifetimes, altruism, and strategic behavior towards heirs. Uncertain lifetimes and incomplete insurance markets result in involuntary bequests as individuals need to save for precautionary motives. If insurance and capital markets are imperfect, uninsured risks related to health and longevity may give rise to precautionary motives for preserving wealth in the old age.

Altruism towards heirs extends the lifecycle savings model to a dynastic framework. When households care about the level of utility of their heirs (and therefore of all future heirs) there is an additional factor that explains increased savings when income increases. Hence, an altruistic retiree would tend not to dissave as quickly as the lifecycle model would predict because of bequest motives. However, the intertemporal allocation of consumption and savings across the members of the dynasty is now done over an infinite time horizon and the intergenerational transfers imposed on households by, say, the introduction of a PAYG pension are completely offset by private behavior and changes in bequests according to the Ricardian equivalence theorem.

Strategic behavior can also justify the bequest motive. Bernheim et al. (1985) argue that bequests can actually be used by self-motivated parents to elicit specific services from their children. Similarly, Kotlikoff and Spivak (1981) argue that the extended family is a substitute for annuities and other insurance mechanisms. Bequests and intravivos transfers may have nothing to do with altruism but are the consequence of

---

27 While the AEW actually increases with low levels of risk aversion, it actually decreases for high levels of risk aversion; i.e., there is a “substitution” effect that dominates.
efficient intergenerational risk sharing arrangements in the presence of incomplete insurance markets. According to the authors, by promising old age support, children are in practice selling an annuity to their parents, the price of which is the bequest.

Empirical studies on the savings behavior of the elderly are central to the explanation of the role of bequests in annuity demand. The life cycle hypothesis predicts rapid dissaving of assets during retirement. However, Bernheim (1987) argues that while bequeathable assets tend to decrease after retirement, more general measures of wealth that include also annuitized wealth (from pensions) tend to remain constant. Annuited wealth can be used to compensate the decrease in bequeathable assets thus supporting the assumption of the bequest motive. Indeed, Bernheim (1991) presents empirical evidence that savings during retirement age is affected by the wish to leave bequests. Hurd (1992) concludes that it is not necessary to invoke the bequest motive to justify slow dissaving during retirement age. In fact dissaving during old age is found to be similar among parents and non-parents providing no support for the bequest motive. The author argues that the absence of private annuity markets can explain slow dissaving in the face of longevity risk.

Laitner and Juster (1996) find substantial evidence of intergenerational altruism in a sample of TIAA-CREF annuitants, which represent relatively high income households. Kotlikoff and Summers (1981) also find high evidence of bequest motivated savings. However, Hurd (1986 and 1989) finds little evidence of intergenerational altruism. Altonji *et al.* (1992) also test the relevance of the altruistic model that predicts that the distribution of consumption among different generations in the dynasty should be independent of the distribution of wealth. By using the panel study of income dynamics on parents and their children in the US, the authors strongly reject the prediction of the altruistic model. Despite observing that a selection bias may exist in the data as wealthier families (more likely to be altruistic and leave bequests) are excluded from the database, they conclude that altruism does not affect consumption, and hence savings, during the old age. However, similar conclusions regarding the relevance of the altruistic model in explaining bequests motives are reached by Wilhelm (1996) by using the Estate Income Tax Match database. This covers only the wealthiest households in the United States and therefore excludes those households unlikely to be altruistic.

Finally, two papers, that attempt a direct relaxation of the Yaari (1965) assumptions are here considered.

Petrova (2002) considers the possibilities that status dependent utility may affect the demand for annuity and that individuals systematically overestimate their life expectancy. The author, using the Health Retirement Study for the US, finds that

---

28 Notice that this is not a rejection of the hypothesis of altruistic transfers among households but only a rejection that the altruistic motives affect savings behavior and hence bequests during old age. In fact, altruistic transfers can take the form of intravivos transfers, like support for education and other in kind transfers earlier in the life of heirs, aimed at offsetting liquidity constraints among heirs (Cox (1990) and Cox and Jappelli (1990)). However, Altonji *et al.* (1997) focus on the more general altruistic assumption and do not find any evidence of altruism in the PSID data.

29 Petrova’s (2002) argument being that, given asymmetries of information between annuity providers and buyers regarding lifespan, a positive bias in individuals’ subjective life expectancy would increase private annuity demand.
contrary to expectations, parental wealth prior to retirement\textsuperscript{30} significantly decreases demand for annuities. Similarly, her measure of overconfidence about individual’s lifespan based on the difference between subjective and observed survival probabilities does not significantly affect demand for annuities.

Davidoff et al. (2003) find more general conditions under which Yaari’s (1965) prediction still hold true. In their paper they find that, for full annuitization to be optimal, sufficient conditions are that 1) annuities pay a rate of return higher than conventional assets net of administrative costs; 2) that capital and insurance markets are complete; and 3) that annuitants have no bequest motives. Partial annuitization is optimal when the condition of complete insurance market is relaxed. The argument being that annuities are “superior” assets than traditional assets as through the pooling premium they pay a higher rate of return.\textsuperscript{31} If only few types of annuities are available and consumers cannot trade with these assets to span all possible states of the world, then consumers need to hold traditional assets that pay a return in those states of the world where annuities are missing or cannot be replicated. The marginal utility from annuitization in the presence of incomplete insurance markets can be negative for higher levels of risk aversion. In fact, the higher the degree of risk aversion the higher the utility derived from conventional assets with positive pay off in those state of the world when annuities do not exist. Hence, as already discussed in section I.C, the monotonic relationship between measures of utility from annuitization and risk aversion is broken when insurance markets are incomplete. Obviously, when also markets for traditional assets are incomplete, it is very well possible that zero annuitization is optimal. Davidoff et al. (2003), however, find this case unlikely and argue that annuitization, higher than the levels observed, should be optimal for most consumers.

II Annuity Supply Constraints

The literature surveyed in the previous section seems to indicate that the observed discrepancy between the prediction of the lifecycle model and the observed small size of voluntary annuity markets could be justified if either consumers have strong bequest motives or if providers charge to high costs for insuring against longevity risk. In either or both cases annuities would not “dominate” (à la Davidoff et al. (2003)) traditional assets anymore. Due to the conceptual difficulty in measuring the bequest motive, or more generally individual altruism, it is likely that appropriate regulation of products, in terms of design, availability and distribution, is likely to be the key determinant for promoting the development of these markets.

In the reminder of the paper we speculate that another reason why annuities fail to dominate traditional assets can be related to excessive risk born by providers of traditional annuity guarantees for a given institutional environment. While the literature presented in the previous section focused on demand constraints on the part of consumers,

\footnote{The author’s hypothesis is that parental wealth should increase demand for annuities as wealth increase individual discount rate.}

\footnote{Indeed the authors argue that any traditional assets is dominated by the same assets with attached longevity insurance if such financial engineering is not associated with too high administrative costs.}
this section focuses on supply constraints on the part of providers.\textsuperscript{32} We briefly summarize the risks that providers generally insure against when selling annuities; we then present a conceptual framework for sharing these risks between providers and annuitants\textsuperscript{33}.

\textbf{II.A A Paradigm of Risk}

Risk, represented by volatility or by the potential for difference in outcome, exists within the annuity environment in several forms and from several perspectives. For sake of simplicity we focus here on longevity or survivorship risk, and investment risk.

First, consider the position of an individual with an accumulated stock of financial assets combined with other resources usually represented by a limited future potential for further earnings, entitlements to social protection from governments and other support from their own networks and family. Exposure to risk can be characterized as the inability to meet consumption needs. This can arise because individuals outlive their available assets and other resources, financial or otherwise, or because the circumstances of their lives change such that their resources prove inflexible or insufficient to respond to this change, or because financial markets fail to perform at the level necessary to deliver on a plan they may have. The plan may be either explicit or implicit. In addition, the alternative sources of support may be identified as inadequate or the social protection or other network contribution may fail – these aspects are largely outside the consideration of this paper.

In terms of mortality, the individual is at risk of what may be considered undue survivorship. The purchase of a life annuity can transfer this risk to the annuity provider but usually brings with it other risks to the individual.

The annuitant is then exposed to the credit risk of the annuity provider. That is, there is a risk that the annuity provider will fail to live up to their obligations. This risk can be reduced through diversification – selecting a number of providers – although this may be limited by the practical constraints such as minimum sizes of investment in the market or limited numbers of available providers. Legislative requirements may also restrict the availability of this option by requiring particular selections or by imposing administrative burdens on individuals which expand with the addition of providers.

The annuitant could also, at least in theory, consider the financial soundness of the provider at the time of purchase. This is, however, generally considered to be difficult if not impossible. Efforts to expose participants to market scrutiny are usually applied; however, the consideration is that they are insufficient so there is also a common policy response to require providers to be subject to additional prudential supervision.

\textsuperscript{32} The distinction between supply and demand constraints is simply dictated by expositional needs. We do not want to imply that there is an excess demand or supply that can be cleared through increased availability of information on either side of the market. Again, it is likely that appropriate regulation of products, in terms of design, availability and distribution, be the key determinant for promoting the development of these markets.

\textsuperscript{33} The section draws on Wadsworth et al. (2001).
In many jurisdictions, the annuitant can be protected from credit risk, at least in part, through the ultimate operation of policyholder protection schemes and policyholder priority in the event of the winding up of the annuity provider.

The annuitant is also exposed to liquidity risk. Life annuities, traditionally, restrict the ability of the annuitant to commute their contract so as to trade off future payment streams for current increased access to cash. Increasingly, products are being developed to permit greater flexibility but at a cost. The alternative for the annuitant is usually a sound financial plan which identifies alternative resources to meet such needs should they arise. It may be, however, that structural impediments also exist that will restrict the annuitant from selecting the desired level of annuitization.

Finally, for the annuitant, can be exposed to investment related risk, even after it has access to annuities. Whilst the annuity provider may have provided a guarantee as to the rate of return implicit in the life annuity, the annuitant may have limited its option to access upside in investment markets after taking out the product. Substantial investment risk is retained in the event that an annuity is not purchased on the downside. The guarantee from the annuity provider is, therefore, a valuable one. But the cost of the guarantee includes a lack of participation in the upside.

Annuity providers have an advantage, relative to individuals, in taking on these risks. They can diversify some of the risks accepted so as to offer value to the customer. This opportunity for diversification exists in mortality and investment related risks.

II.A.1 Mortality risk

With respect to mortality, the risk that one individual may have greater survivorship than another can be diversified through the pooling of risk from many annuitants. This pooling operates whether the market has a degree of imposed compulsion or not. However, levels of expected mortality will vary reflecting the extent of selection available to the individuals. The risk to the annuity provider in this respect is that it may misestimate the general level of mortality that would be applicable to the annuity pool – a systemic risk, perhaps considered as model risk or pricing risk. This risk can be mitigated through the collection of appropriate data and the sound analysis and study of this data and the experience. The remaining risk for the provider can be that it has a portfolio of insufficient size for the effects of the mortality pooling to take full effect. In this case considerations towards reinsuring such “excess” risk or increasing capital to cover the prospect of adverse outcomes need to be taken.

Mortality for a group of annuitants is, however, not quite as predictable as might be desired. There is a systemic element represented by the general trend of mortality improvement that is observed in many countries. This can also be considered to be a pricing or model risk; however, being systemic, it is not diversifiable as it applies to the pool as a whole.

Predicting mortality improvement remains problematic. In some cases, mortality measurements are hampered by data problems. These can be further exacerbated at older ages rendering this of particular interest with respect to retirement income analysis. At most advanced ages, the age pattern of mortality is not as well understood as is desirable. The rate of decline in mortality at these ages is, as a consequence, also unclear. As a
result, efforts to improve data collection and to research the influences on mortality continue with respect to these aspects.

Even where the data is more reliable, the pattern of improvement over time, as observed through the twentieth century, has been variable and its prediction is problematic. The rate of improvement varies from country to country, it is inconsistent over time even within the one country, and it varies by groups of the population such as by age group or sex. Developing an improvement assumption for the future projection of mortality has met with considerable uncertainty. Unfortunately for the providers of annuities, the experience in both developed and developing countries has been that the rate of improvement is most often found to have been underestimated. Underestimation of improvements means that companies have to increase provisions – in effect meaning that capital is the ultimate protection against this risk.34

Companies can respond to this risk by allowing for adequate or even conservative (in this case conservative is equivalent to aggressive expectations) mortality improvement in pricing and reserving. However, the resource to provide protection beyond the even conservative assumptions in the liability valuations will ultimately be capital. Prudential regulation should require liability valuations which are sound and minimum capital requirements that provide an adequate buffer over and above these provisions.

II.A.2 Investment risk

Investment risks to the provider can be diversified through effective investment strategies. Diversification of credit risk in investments is available to annuity providers. Diversification and liquidity can be greater for institutional investors than individuals so will be positive in the reduction of risk. Structural rules can, however, limit the available diversification either through investment rule restrictions or a lack of liquidity or diversification available in investment markets. Alternative investment markets may also introduce currency risk which can be reduced through hedging instruments if available. It is rare that such hedging instruments would be readily available at the durations necessary with appropriate liquidity so some element of risk will be retained.

Investment risk cannot be adequately diversified by the annuity provider as the liabilities to the annuitant pool will generally be of a duration longer than the available assets to match this risk. The result is that investments will mature and need to be reinvested at uncertain rates of return before they are ultimately needed to make payments to customers. Providers usually respond to this risk through the holding of capital and the lobbying for longer term investment assets. In some cases, the counter-party may well be interested in providing such long term instruments as it has similar long term requirements – in effect this is the access of diversification through the capital and debt markets.

In other cases, the provider may be motivated by other concerns. This is likely in the case that governments provide such instruments. Where there are no pure financial objectives for the provider of longer term instruments to issue them then the capital

---

34 To a limited extent, reinsurance may also provide protection however this exposes the provider to the credit risk of the reinsurer.
requirement of the annuity provider is, in effect, reduced and transferred to the debt instrument issuer.

By way of observation, mortality change is essentially identical in effect to a change in the rate of interest guaranteed in the annuity and yet the amount of concern expressed to seek interest rate matched products would seem to need to be matched by efforts to ensure that mortality improvement is not misstated. Most often, however, it is given greater concern perhaps as it is considered more amenable to change. As interest rate structures move to structurally lower nominal levels, as they have in many jurisdictions, the potential for the effects of misestimated mortality is likely to be of greater financial importance than the effects of interest rate mismatching.

Capital, as the solution to these risks, for the annuity provider is also subject to operational type risks. We have indicated that capital is of use to protect the annuity provider, and ultimately the annuitant through the credit risk associated with the provider, from pricing risks, market risks, and mortality risks both through the risk of small pools or systemic mortality improvement. Capital also provides ultimate protection against operational risk in the annuity provider. Where annuity providers consider only economic capital and this encompasses all of these risks then there is no real distortion. Where, however, companies focus on statutory capital requirements then there can be a distortion in their behavior.

Ultimately, society and governments can be exposed to the risks of annuity markets through these mechanisms. If the government mandates annuitization then it has a stake in the performance of the annuity providers. If the government issues long term assets and the capital requirements are reduced as a result of companies accessing improved matching of assets and liabilities then the performance of those long term assets becomes critical – in effect the capital requirement of the annuity provider may be transferred to the government. Finally, there is the implied or explicit response to failure of an annuity provider.

An alternative is to consider the sharing of these various risks between the state, providers and individuals, as it is explored in the subsequent sections of this paper.

**II.B Constraints on annuity supply**

In this section we present a conceptual framework for sharing longevity and investment risk between providers and annuitants. The following figure depicts a continuum of sharing arrangements that can be reached by these two agents from the point of view of the consumer.
On the vertical axis we measure investment risk while on the horizontal axis we measure longevity risk. A point like A would correspond to a guaranteed lifetime annuity where both investment and longevity risk are borne by the provider. The consumer in A bears no risk. A point like B would correspond to the situation of self-insurance. For instance, the consumer receives a lump sum only from its pension assets and bears full investment and longevity risk. In B there is no access to annuity the market. A point like D would correspond to the purchase of a variable, or with profit, annuity. Here the consumer bears the investment risk but no longevity risk. A point like E would correspond to a product the income from which varies with survival experience but investment return is guaranteed. Finally, a point like C would correspond to a product with income (investment and survival rates) reviewable at regular intervals.  

The pension regimes of many countries tend to promote structures at the extremes represented by A and B. Either investment and mortality guarantees are carried largely by a provider (an insurance company or pension scheme) or pensioners bear all of the investment and survival risk associated with provision of a lifetime income, for example, making withdrawals from an investment fund or bank account. In this second situation of self-insurance, the pensioner may be bearing investment and survival risks that subsequently prove to be too heavy. 

The exposure to decline in income faced by a pensioner relying on self-insurance relatively to a situation where longevity insurance is provided for can be summarized as follows. Let’s assume that the pensioner anticipates a future annual investment return of

---

35 For instance, TIAA-CREF in the US or Prudential in the UK offers this type of annuity. ABP and a number of Dutch pension schemes are exploring moving to risk sharing models. Danish pension schemes already use risk-sharing models in the payment phase as do Norwegian ones.
i and plans to take a fixed stream of pension withdrawals each year for life. The actual return to be earned on the retirement fund simply in order to sustain the target income can be represented by \[ (i + \frac{q_x}{1-q_x}) \] where \( q_x \in (0,1) \) is the probability of an individual dying aged \( x \); i.e., between \( x \) and \( x+1 \).

At early post retirement ages, \( q_x \) is typically relatively small. Provided the pensioner has been cautious in his anticipation of future investment returns so that the return actually achieved \( i' \), is larger than \( i \), he may find that \( (i'-i) > \frac{q_x}{1-q_x} \). However, sustaining this outcome becomes increasingly unlikely as \( q \) rises with age; i.e., \( x \). The following table illustrates what additional rates of return (in the second column) should be required by the self insurance strategy for increasing probabilities of dying (in the first column)

**Table 2: Self insurance strategy and required rates of return**

<table>
<thead>
<tr>
<th>( q_x )</th>
<th>( \frac{q_x}{1-q_x} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>4.20</td>
</tr>
<tr>
<td>0.10</td>
<td>11.10</td>
</tr>
<tr>
<td>0.25</td>
<td>33.30</td>
</tr>
</tbody>
</table>

*Note: \( q_x = 0.01 \) means the chance of dying between year \( x \) and \( x+1 \), for a person aged \( x \), is 1 in 100.*

We may refer to the additional return required, i.e., \( \frac{q_x}{1-q_x} \), as “the cost of survival” for a retiree relying on self insurance. In the OECD countries, the \( q_x \) listed in the first column of Table 2 may be broadly related to mortality rates of individuals in their sixties, seventies, eighties and nineties, respectively.

Self-insuring survival risk has traditionally been accompanied by greater freedom of investment than applied to annuities. This freedom of investment might in principle be used to generate a higher investment return (accompanied by higher risks and the possibility of a worse outcome). However, the “cost of survival” at later ages is higher than for example the equity risk premium typically identified for major markets.

Pooling survival risk with other lives provides a more effective hedging mechanism for meeting some or all of the “cost of survival” since the resource transfer from those that die to those that survive also takes the form \( \frac{q_x}{1-q_x} \), but may be regarded as unsatisfactory by a pensioner if investment choice is restricted.

Annuities provided by insurance companies are typically positioned at the other extreme at (or close to) A in Figure 1. The provider guarantees in full both investment returns and survival rates and a large capital backing for these annuity promises is likely to be needed. Such capital requirement should be a an increasing function of: 1) the mix

---
36 See APPENDIX A.
of risk exposures retained by the provider; and 2) the level of security sought. In what follows we explore the interaction between annuity promises, risk and capital requirements, focusing on the simple example of a provider supplying a level income life annuity to a retiree aged 60 (with a life expectation of around 20 years). For the sake of simplicity we imagine the provider operating in 3 notional countries: I, II and III. Although the characteristics set out below for the bond markets in the 3 notional countries are illustrative and not precisely representative of actual countries or markets, the markets may be regarded respectively as a relatively well developed and liquid market with a broad range of bond offerings, a more limited but still mature market focused on shorter term debt and a similarly limited but immature and more volatile market. The key features of the bond markets are for our purposes:

- **Country I**: a plentiful supply of bonds of all durations with a yield curve as set out in 0.
- **Country II**: no bonds of duration greater than 5 years and an initial yield curve as set out in 0. In this case, as most annuitants will live longer than 5 years, funds to support continued annuity payments will need to be invested on unknown terms at the end of each 5 years. A distribution of interest rates available at each reinvestment date is assumed (as set out in APPENDIX C). The annuity provider runs the risk that interest rates on reinvestment will be lower than assumed in setting the annuity rate and the variation could be substantial.
- **Country III**: the yield curve is as for Country II and no bonds are available for durations greater than 5 years. However, as the assumed bond market is less well developed than in Country II, the distribution of interest rates for reinvestment after 5 years is wider -volatility is higher by 50%.

For our examples, the survival risk we assume relates only to the systemic error referred to in II.A.1 above as follows:

- **Country I**: survival rates have generally developed in line with estimates.
- **Country II**: there has been a tendency for survival rates to improve faster than estimated and it is felt on medical and other grounds that there is a significant possibility of future improvement above current best estimates. Reduction in death rates above best estimates may be about 1% pa and this “error” equates for a 60 year old annuitant to about 0.25% pa of lost investment return.
- **Country III**: survival rates have also tended to be underestimated and the data is felt to be even less reliable. Possible reduction in death rates could be 2% a year equivalent to about 0.5% pa of lost investment return.

For Country I we are confident of the future rates of interest and longevity and calculate our annuity rate accordingly. For Countries II and III, future interest and survival rates are uncertain. For any given level of annuity payment, we can estimate the

37 There is considerable academic debate over potential variation in future rates of longevity improvement, even when substantial data exists to determine recent past experience – see Oeppen et al. (2002) and Olshansky et al. (2001) for a discussion.
possibility that the initial annuity fund plus any capital backing plus interest less payments and expenses will be exhausted before the annuitant dies. The annuity provider (and the regulatory regime in the relevant country) will usually be cautious – that is allow for a relatively low possibility of not being able to make the annuity payments. We have assumed in our example a maximum failure rate of 1%. The possibility of failure can be changed either by varying the annuity rate or the capital available to support the annuity.

The results from our example are presented in the table below. The table shows the contribution made by each of short bond durations, volatility of interest rates and mortality uncertainty to either reducing annuity rates or requiring additional capital.

Table 3: Comparison of annuity rates or capital requirements between countries

<table>
<thead>
<tr>
<th></th>
<th>Bonds less than 5 years</th>
<th>50% increase in volatility</th>
<th>Mortality uncertainty</th>
<th>Total adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduction in annuity rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country II</td>
<td>-8%</td>
<td>-</td>
<td>-2%</td>
<td>-10%</td>
</tr>
<tr>
<td>Country III</td>
<td>-8%</td>
<td>-3%</td>
<td>-4%</td>
<td>-15%</td>
</tr>
<tr>
<td>compared with Country I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increase in capital</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country II</td>
<td>+9%</td>
<td>-</td>
<td>+2%</td>
<td>+11%</td>
</tr>
<tr>
<td>Country III</td>
<td>+9%</td>
<td>+4%</td>
<td>+5%</td>
<td>+18%</td>
</tr>
<tr>
<td>compared with Country I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By way of illustrating the application of the numbers in the table above, let us assume an annuity purchased for 100,000 local currency units (LCUs) in Country I delivers an income each year of 7,000 LCUs. The annuity deliverable (with 1% risk of failure) for 100,000 LCUs in Country II is 7,000 x .9 = 6,300 LCUs or in Country III is 7,000 x .85 = 5,950 LCUs. The reduction in annuity between Countries II and III can be attributed to increased uncertainty (volatility) of interest rates on reinvestment (= 7,000 x .03 = 210) and greater mortality uncertainty (= 7,000 x .02 = 140). Alternatively, an annuity provider may give the same annuity rate for all 3 countries but provide additional capital to ensure no more than 1% risk of failure in Countries II and III. Additional capital required in Country II is 100,000 x .11 = 11,000 LCUs and in Country III is 100,000 x .18 = 18,000 LCUs. Additional capital required between Countries II and III reflects increased uncertainty (volatility) of interest rates (= 100,000 x .04 = 4,000) and greater mortality uncertainty (= 100,000 x .03 = 3,000).

The example necessarily oversimplifies the situation likely to apply for any given country and provider. However, lack of suitable matching assets and underestimation of longevity improvement either singly or in combination are likely to be applicable to many countries that will as a consequence of an ageing population be considering requirements for a retirement income structure.

**II.C Alternative solutions**

The examples in the previous section illustrate the argument that guaranteed lifetime income (as at position A in Figure 1) can require substantial capital backing by the
provider and/or be very expensive; i.e., low capital into income conversion rate, for the pensioner. The regulatory implications of this are obvious: mandating guaranteed provision in excess of what can be fairly and profitably delivered by private providers may result in inefficient allocation of resources. Also, capital may not be available in the quantities required to meet the demands of our ageing population. However, total self insurance (as at position B) may also, as we have seen, be unattractive.

The alternatives shown by A and B in Figure 1 represent extreme solutions, either of which may be sub-optimal for the pensioner. The optimal level of guarantee for a pensioner can be expected to vary with the price attached to the guarantees and the value perceived to attach to the implied security. It might for example be located at position C in Figure 1 with the annuitant sharing some investment and some mortality risk with the annuity provider. In Countries II or III above, a response by providers to bond supply limited to a maximum 5 year duration and uncertainty as to the rate of longevity improvement could be to offer a reviewable annuity, for which at, for example, 5 yearly intervals, ongoing income payments were reviewed and if necessary reset to reflect current interest and survival rates.

If for example the supply of long duration price index-linked bonds in a country was small or non-existent, the terms on which price index-linked annuities could be provided in volume to pensioners may be relatively unattractive compared with, for example, a variable interest rate product where the level of income payments was reviewed at regular intervals and where interest rates were (broadly) correlated with levels of inflation. Such a solution is located at position D in Figure 1.

In position E, rates of interest may be guaranteed for life but annuity terms would be reviewable to reflect changing longevity experience. An annuity which guaranteed lifetime income regardless of improvements in survival rates after it was purchased may contain a substantial margin for uncertainty in its pricing (or alternatively if such a margin is not included carry a default risk for the annuitant). A willingness by the annuitant to accept regular reviews could result in the provider stripping out most of this margin.

An approach which allows at least some intra generational exposure to survival risk may also allow some pooling of survival risk to be maintained; the alternative may be withdrawal of annuity providers from survival pooling. The choices can be illustrated as follows:
Obviously, the return from pooling could be higher the higher the level of guarantee. A schedule like 1 in Figure 2 would be typical of a product like A (guaranteed annuity) in Figure 1. This would allow for expected mortality plus a margin for the guarantee against uncertainty and any associated costs of additional capital. A range of outcomes like 2 in Figure 2 could be typical of a product like C (some risk sharing between provider and retiree) in Figure 1 with 2'' being the “best estimate” schedule at retirement. In this case costs of improvements in survival rates above those expected at retirement are borne by pensioners and could produce an outcome as shown in 2''''. However, an outcome as in schedule 2' would also be possible. These are respectively less and more favorable for the pensioner than the guaranteed outcome. A schedule like 3, along the horizontal axis, is typical of a self insurance case B in Figure 1; for B, the pooling premium is obviously zero.

As previously mentioned, the pricing of uncertainty in survival rates is not well developed, in particular uncertainty as to future improvement, and it may be that the margins charged for uncertainty will in future increase. This might, for example, be a consequence of developing benchmark prices through a market in longevity bonds. The reinsurance market for longevity risk could also fulfill this role, but is currently small and immature. In countries in which annuity volumes could grow rapidly as a consequence of an ageing population, self-financing of capital by insurers may not be possible and capital markets would have to be persuaded that reasonable returns relative to risk could be earned on writing annuity business.

As partly discussed in section I.D the willingness of pensioners to self-finance some survival risk might be expected to be higher, if the utility from bequests is high, or if individuals are member of a “self-insuring social group”, or when they are at (relatively younger) ages at which the returns from pooling are low, or at (relatively younger) ages
at which it remains possible/desirable to work, or when guaranteed pooling involves loss of control over asset allocation or the pacing of income withdrawal.

The evidence from many countries (see Section I.B) is that where survival risk pooling is voluntary, it is widely rejected at or around normal retirement ages. However, the reverse might well tend to apply at older ages, where the rewards for pooling survival risk are high.

In summary, based on Figure 1, we might have a flexible retirement income vehicle which allows a choice of investment and pooling guarantees to pensioners. Pensioners could choose:

- investment exposure from volatile funds to secure or guaranteed returns with scope to vary the mix over time
- survival guarantees from nil to periodic review to guaranteed for life, with scope to increase pooling and guarantees with increasing age
- the rate of withdrawal of “income” from their retirement funds.38

II.C.1 Rating survival probabilities.

One of the disadvantages for some pensioners of annuitization identified in academic studies has been that of selection arising from the pooling of survival. For some groups of pensioners, for example those with health impairments or those with poorer socio-economic backgrounds, the terms on which pooling takes place may mean a high probability of subsidizing other parties to the pool because of the lack of homogeneity of lives. This effect can be seen as reg emissively redistributive. Optimal pooling arguably takes place when pooling relates to sharing risks where the outcome is sufficiently uncertain and the terms for pooling are such as to make hedging the risk attractive.

Rather than a return for pooling at age $x$ being $\frac{q_x}{1-q_x}$, rating and/or underwriting may be applied to the $q$ factors to take into account particular characteristics of the pensioner relevant to anticipated death rates. Thus for pensioner $r$, the pooling benefit at age $x$ may be $\frac{q_r^x}{1-q_r^x}$, where $q_r^x$ may, subject to legal/regulatory constraints, take into account some or all of medical history, health indicators (i.e., body mass index), social indicators, (i.e., education, or past occupation).39 Precise mechanisms need careful thought, but in principle there is scope to achieve pricing of retirement income streams that are seen to be more equitable than a more general pooling of life risk. Those with the poorest survival probabilities will be subject at each age to the highest mortality, i.e., $q_r^x$ factors, and therefore the greatest pooling benefit.

---

38 The complexity of such a model can be tailored to fit the capabilities of providers and the financial system and the sophistication or otherwise of pensioners. A very simple model might, for example, consist solely of interest bearing deposit accounts and survival risk-pooling on half the fund, with income withdrawals reviewable at 5 yearly intervals.

39 Such “multivariate” rating can be found in other forms of insurance such as motor or household.
The introduction of rating would have the impact of eliminating the non-stochastic component (i.e., those elements that would induce strong adverse selection) from the pooling equilibrium. In other words, types systematically better than the average (i.e., who die earlier) would be better off, while types systematically worse than the average would be worse off. This form of separation of types may not be fully acceptable at a society level\(^{40}\). Hence, it may be applied differently to the different sources of retirement income.

**II.D Impact on minimum capital requirements**

The flexible retirement income model could be operated with substantially less capital backing than would be required for traditional lifetime guaranteed annuities and consequently reduced constraints on supply. If guarantees are typically not exercised until later post retirement ages, the cost of the guarantee is proportionate to the then remaining fund, and assets of shorter matching duration are more plentiful, then the combined effects can represent a dramatic reduction in capital requirements. The following example illustrates this point.

Let define with: \(F_x\) the level annuity purchase price at age \(x\); \(CG_x\) the capital needed to back full longevity and investment risk insurance at age \(x\) as percentage of \(F_x\); \(q_x\) the probability of dying between ages \(x\) and \(x + t\); and \(r\) is the required return on capital. Then:

\[
CG_x F_x
\]  

is the capital needed for a provider to fully insure annuitant’s longevity and investment risk at age \(x\) (say, retirement age), in position A in Figure 1.

It is fair to assume that \(CG_x' < 0\) and \(F_x'' < 0\) as the shorter the coverage period of the guarantee: 1) the higher the probability that appropriate maturity matching instruments become available; and 2) the lower the mortality forecast error.\(^{41}\)

Let’s now assume that retirees have only access to longevity and investment guarantees \(t\) years after retirement. In other words, retirees start in position B in Figure 1 at age \(x\), revert to position A at age \(x + t\), while between \(x\) and \(x + t\) retirement income is generated from withdrawals from fund. Then:

\[
CG_{x+t} F_{x+t} (1 - q_x)(1 + r)^{-t}
\]  

is the value at outset of the capital needed for a provider to fully insure annuitant’s longevity and investment risk starting at age \(x + t\) multiplied by the probability that the retiree survives until age \(x + t\). It then follows that:

\(^{40}\) Separation could occur along gender, racial, education, or health lines, for instance.

\(^{41}\) It is likely that the price of the guarantee be also influenced by the return on pooling. This increases with age and should, *ceteris paribus*, have a positive effect on prices. For simplicity, we assume that this positive impact does not offset the negative impact of 1) and 2) above.
\[ CG_x F_s - CG_{x+t} F_{x+t} (1 - q_s) (1 + r)^{-t} = 1 - \frac{CG_{x+t} F_{x+t}}{CG_x F_s} (1 - q_s) (1 + r)^{-t} \]  

(3)

is the reduction, valued at \( x \), in capital requirement for a provider in the two states of the world (A, and B with reversion to A at a later date in Figure 1).

In a real world case where \( x = 65; \) \( (1 - q_{65}) = 0.5 \); the capital requirement to back the annuity (as a percentage of purchase price) falls from 25% at age 65 to 10% at age 85 (i.e., \( CG_{65}/CG_{65} = 0.4 \)); the purchase price for a level annuity at age 85 is 60% of that at age 65; (i.e., \( F_{85}/F_{65} = 0.6 \)); and the cost of capital over the period is \( r = 10\% \); then:

\[ 1 - \frac{CG_{x+t} F_{x+t}}{CG_x F_s} (1 - q_s) (1 + r)^{-t} = 1 - 0.4 \cdot 0.6 \cdot 0.5(1 + 0.1)^{-20} = 0.9822 \]  

(4)

is the reduction in capital requirement achieved in this example by guaranteeing annuity payments from age 85 instead of 65.

**II.E The risk sharing framework for the whole pension system**

So far we have discussed the case of risk sharing arrangements between private sector annuity providers and annuitants. The main argument was that such arrangements could relax supply constraints and help promote the development of annuity markets. However, not all annuities are provided through private sector providers backed by capital. Across countries, a variety of institutional arrangements are in place to provide annuities to retirees.

Private pension schemes, (both state and private sector) in many countries provide income for life that has many of the characteristics of an annuity. The ability of the scheme to bear risk and to guarantee income payments rests on the financial strength and capability of the sponsor rather than on capital, and members of the scheme are therefore exposed to “sponsor” risk. This risk is of course particularly great in public unfunded pay-as-you-go schemes where the guarantee is supported only by intergenerational transfers of wealth and by the fact that the pension promise is dynamically inconsistent. Funding of schemes, depending of the level achieved, will help to reduce to a greater or lesser extent, the exposure to sponsor risk.

Pension scheme structures provided by the state or by employers generally do not appear to charge pensioners for investment or longevity guarantees. The pension guarantee can be softened in a variety of ways which transfer some risk to the pensioner that is moving in Figure 1 from position A in the direction of although not necessarily reaching position C. For example, only a level pension in payment may be guaranteed with any increases being discretionary and dependent on the performance of the associated investment fund (or economy). In addition, lower levels of discretionary increase might be granted where for example survival rates improve faster than anticipated. Additional movement away from guarantees and towards risk sharing may involve periodic adjustment of pensions in payment to reflect associated investment/ economic performance or survival experience. However, in several countries (see examples in Section 5), private sector schemes operate in the top right hand corner of
Figure 1, merely paying out a lump sum at retirement, thereby passing on all investment and survival risks to pensioners.

A major issue for pension schemes is the increasing length of the post retirement period and the scope through technological, medical or other advance for unsupportable longevity improvements to take place. If the terms for retirement income are guaranteed and cannot be varied post retirement, the implication is that longevity guarantees will be underwritten for each retired generation by the successor generation(s). Such a model may come under threat for many reasons including:

- material unexpected improvement in longevity
- lack of will or financial resources to meet the additional longevity costs, for example as a consequence of a declining financing base, whether state (ageing population) or corporate (downsizing).

Alternative outcomes to greater risk sharing may be either or: 1) partial or total default by the scheme or sponsor; 2) reducing future accrual of pension guarantees; 3) withdrawal from ongoing pension provision. However, the alternative to greater risk sharing may be providers of all types (pension schemes and insurers) withdrawing as suppliers or only offering retirement income structures that place the risks with pensioners. Members of pension schemes may also be exposed to the risk of ‘sponsor’ failure.

It may be useful for policymakers to view the relationship between the different elements of a pension system in terms of a spectrum of risk sharing possibilities in relation to lifetime income (Shiller 2003). In other words, countries’ pension systems should be designed in a multipillar fashion with increasing pillars or layers, providing a decreasing level of income guarantee (Holzmann et al. 2003) to mimic the decreasing risk aversion that stems from higher income levels. For instance, a social safety net financed through general taxation could provide full guarantee to the indigent who has not been able to participate to the formal pension system. A publicly managed pay-as-you-go pillar financed through intergenerational transfers could provide full basic guarantee to all participants in the system. Higher pillars in the system, to which more affluent people would participate if non-mandatory, would provide increasing risk transfer from the society to the individual. Lindbeck and Persson (2003) reach a similar conclusion when discussing how different elements of a pension system provide different means to redistribute risk intergenerationally, intragenerationally and along the life cycle. According to the authors, annuitization risk is a special type of intergenerational risk composed of two essential element: 1) a “transformation” risk related to the specific condition of the capital market in which a given cohort retires and that would determine the lifetime income of that cohort; and 2) a longevity risk that could be shared between providers and annuitants. Due to these and other risk sharing characteristics discussed in the paper Lindbeck and Persson (2003) conclude that “It is, in general, advantageous to combine funded and pay-as-you-go systems, since they have different risk characteristics, with respect to both market risk and political risk. It is probably also a good idea to combine a defined contribution system with some elements of a defined benefit system” (page 104) as these elements redistribute risk differently among and within generations.
III CONCLUSIONS

The theoretical literature on wealth allocation, based on the life-cycle hypothesis, predicts that individuals would be better off by holding only annuitized assets in the absence of a bequest motive, or a portfolio of annuitized and traditional assets in the presence of a bequest motive. The strong assumptions of earlier literature have been relaxed and more recent literature suggests that the “superiority” of annuities depends on: 1) limited importance of the bequest motive; 2) completeness of the insurance and capital markets; and 3) annuities paying a higher rate of return (net of administrative costs) than traditional assets. The “superiority” of annuitized assets is also supported by the results of the literature on the money’s worth of annuities.

However, factual evidence shows that annuity markets, especially voluntary annuity markets, appear to be very small in most countries: i.e., annuities represent a less than expected share of household wealth. Empirical studies on the bequest motive provide ambiguous results on the importance of such a motive to hold traditional assets and it is likely that the incompleteness of insurance markets is the most important cause for the perceived “over investment” in traditional assets. When insurance markets are incomplete, individuals need to save in traditional assets with a positive payoff in those states of the world where insurance cannot be obtained.

In addition, the small size of voluntary annuity markets may be related to traditional products that seem to provide consumers with only a binomial allocation of risk. Either self-insurance or complete insurance of the most important types of risks are available in many countries. Lifetime guarantees of income effective from the date of retirement may be very expensive (if backed by capital as in the case of insurance companies) or result in low annuity rates or be vulnerable to default in the case of pension schemes. Also, pooling of survival risk may appear unattractive at typical retirement ages, but valuable at later ages. As the size of the retired population in many countries increases and the length of time spent in retirement extends for the average individual, pressure to maximize income from residual assets is likely to grow.

An unbundled model for retirement income opens up the possibility for greater risk sharing between pensioners and pension providers. Risks to be shared relate in particular to investment returns (including in principle returns such as growth of national income) and survival rates. Long-term improvements in survival rates are difficult to forecast or to hedge, and may in particular be shared with or retained by the annuitant population. Optimal risk sharing, for example the duration and extent of guarantees, will depend on the pricing of any guarantees and pensioners’ willingness to pay the price (or accept the risk of default). Unbundling of survival risk can be extended to differentiating the terms for pooling risk so that those expected to have poorer survival rates receive higher income. This can overcome or at least reduce regressive adverse selection effects, but the extent and nature of such differentiation to be allowed requires careful consideration by policymakers.

The rationale for risk sharing applies to the whole pension system in any country. So far, the excessive focus of the policy debate on pension reforms on accumulation has left many countries with inadequate retirement products. In these countries, a review of the regulation of retirement products is required before too many new beneficiaries retire.
from the system and the political costs make changes unfeasible. Ideally, the level of longevity and investment risk guarantee could be means tested and be a function of the development of domestic financial markets.
REFERENCES


A Conceptual Framework For Retirement Products: Risk Sharing Arrangements Between Providers And Retirees


APPENDIX A

Let:

Retirement income fund at time $t = F_t$

Annuity payments set at time $t = \frac{F_t}{\overline{a}_x^i} = A_t$

where $\overline{a}_x^i$ assumes annuity payments of 1 per period to lives aged $x$ at the start of the period, and anticipates an investment return of $i$ p.a.

The fund $F_{t+1}$ required for survivors from this group at $t+1$ in order to maintain annuity payments of $A_t = A_t \overline{a}_{x+1}^i$

$$\overline{a}_x^i = (1 - \frac{q_x}{2})\left(\frac{1}{1 + \frac{i}{2}}\right) + (1 - q_x)(1/(1+i))\overline{a}_{x+1}^i \quad \text{(Definition)}$$

Note: this approximates 'continuous' payment of the annuity over the period with a payment made at the mid point in the period.

Therefore, $\overline{a}_{x+1}^i = (1 + \frac{q_x}{1 - q_x})\left[\overline{a}_x^i (1+i) - (1 - \frac{q_x}{2}) (1 + \frac{i}{2})\right]$.

Let $(1 - q_x) = 1/(1 + i')$

Then $i' = \frac{q_x}{1 - q_x}$

Note: $0 < 1 - q_x < 1$

Then $\overline{a}_{x+1}^i = \overline{a}_x^i (1+i') (1+i) - \left(1 + \frac{i}{2}\right)\left(1 + \frac{i'}{2}\right)$

Multiplying by $A_t$

Then $F_{t+1} = F_t (1 + i + \frac{q_x}{1 - q_x}) - A_t \left[1 + (i + \frac{q_x}{1 - q_x})/2\right]$.

Note:

If annuity payment is at start year:
\[ \frac{F_t}{(F_t - A_t)(1 + i + \frac{q_x}{1 - q_x})} \]

If annuity payment is at end year:

\[ F_{t+1} = F_t (1 + i + \frac{q_x}{1 - q_x}) - A_t \]
Figure 3: Yield curve for Country I

Figure 4: Yield curve for Country II and III
Assumed frequency distribution of the 5 year yields at the end of 5 years for Country II

Note: the distribution of interest rates was derived by calibrating a financial model using swaption data consistent with the yield curves in Appendix B