# The Liquidity Premium for Illiquid Annuities

S. Browne,
Columbia Business School
Columbia University
3022 Broadway, Uris Hall 402
New York, NY 10027

M. Milevsky,* T. Salisbury
York University
4700 Keele Street
Toronto, Ontario
M3J 1P3, Canada

October 15, 2002

*Moshe Milevsky, the contact author, can be reached at Tel: 416-736-2100 x 66014, Fax: (416) 736-5487, or via Email milevsky@yorku.ca. This research is partially supported by the Mathematics of Information Technology and Complex Systems (MITACS) as well as The IFID Centre at the Fields Institute. Salisbury acknowledges support from NSERC. The authors would like to acknowledge helpful comments from M. Cao, P. Chen, G. Daily, J. Francis, A. Gottesman, G. Jacoby, S. Posner, K. SigRist and two anonymous JRI reviewers. A previous version of this paper circulated under the title of: *A Note on the Solid Value of Liquid Utility.*

# The Liquidity Premium for Illiquid Annuities.


## ABSTRACT

Academics and practitioners alike have developed numerous techniques for benchmarking investment returns to properly account for excessive levels of risk. The same, however, can not be said for liquidity, or the lack thereof. This paper develops a model for analyzing the *ex ante* liquidity premium demanded by the holder of an illiquid annuity. In the U.S., an annuity is an insurance product that is akin to a pension savings account with both an accumulation and decumulation phase.

We compute the yield (spread) needed to compensate for the utility welfare loss, which is induced by the inability to rebalance and maintain an optimal portfolio when holding an annuity.

Our analysis goes beyond the current literature, by focusing on the interaction between time horizon (both deterministic and stochastic), risk aversion and pre-existing portfolio holdings. More specifically, we derive a negative relationship between a greater level of individual risk aversion and the demanded liquidity premium. We also confirm that, *ceteris paribus*, the required liquidity premium is an increasing function of the holding period restriction, the subjective return from the market, and is quite sensitive to the individuals endowed (pre existing) portfolio.

"..*If the insurance company has greater and longer surrender charges, then it can pay more, on fixed annuities, knowing the funds aren't going to leave, so the liability structure will be more stable*...**Best's Review**, October 2001, pg. 43"

# 1 Introduction and Motivation

In the United States, the term annuity covers a wide spectrum of financial and insurance products. A *savings* (pay-in) annuity is akin to a bank account or savings bond where money is accumulated over a period of time at a variable or fixed rate of interest. In contrast, a *consumption* (pay-out) annuity is similar to a pension that pays a periodic fixed or variable amount, which might also contain longevity insurance. The former is usually used prior to retirement, while the later is used during the retirement years. According to the above-quoted A.M. Best survey, more than $1 (U.S.) Trillion is currently invested in various types of annuity products in the U.S.

The common denominator of fixed (in contrast to variable) *savings* and *consumption* annuities is that they are quite illiquid. Namely, in stark contrast to a money market fund or savings bond that can be redeemed on a daily basis without any penalty, it is difficult or very costly to surrender (or cash-in) a fixed annuity. While the reasons for this illiquidity differ depending on whether the product is in the accumulation or decumulation phase, the fact remains that continuous asset re-allocation is virtually impossible with these products. Our paper therefore asks a simple question: *What is the liquidity premium that a rational investor will demand to compensate for the illiquidity?*

Thus, for example, the holder of a fixed (savings) annuity might be told that he or she cannot withdraw from (or cashout of) the product for the first seven years of the contract. Or, in the event of a permissible early withdrawal during first seven years, one might be 'hit' with an $x\%$ penalty, a.k.a. market value adjustment. In the payout phase the liquidity restrictions can be even more severe. And, while in the decumulation phase, the illiquidity is often accepted as the cost of obtaining longevity insurance, in the accumulation phase *policyholders are told* that their return will exceed the yield of a comparably liquid instrument. Indeed, it is quite common to see a monotonic relationship between the magnitude of the early surrender charges on a fixed annuity – controlling for commissions – and the guaranteed yield *if* the product is held to maturity. Implicitly, investors (or more precisely, policyholders) are being promised compensation for the liquidity restrictions.

Recent academic literature has documented the empirical welfare gains from annuity products and annuitization, as well as the value of longevity insurance. For example Mitchell et. al. (1999) argued that consumers would be willing to 'give up' to 30% of their wealth to obtain a fairly priced annuity. Likewise, Brown and Poterba (2000) explained the extremely low levels of annuitization, by arguing that married couples function as a mini annuity market. Blake and Burrows (2001) focused on the undesirable longevity risk taken by insurance companies issuing payout annuities, and the need for governments to issue mortality-linked bonds.

However, most of the literature discussing the costs and benefits of annuitization, has ignored some of the problems created by having a portfolio that cannot be liquidated or rebalanced for long periods of time. This is not just an issue in the (relatively unpopular) payout phase of a life annuity, since the same types of restrictions apply in the accumulation phase as well.

Indeed, most financial economists would agree that one should be compensated by the insurance company for the illiquidity restrictions. In other words, all else being equal, a fixed income instrument that cannot be sold – or, for that matter, subsequently repurchased – over the life of the product, should provide investors with a higher yield.[1]

Note, of course, that from the insurance companies' perspective, these restrictions are absolutely necessary to manage the duration mismatch (or risks) that otherwise would arise if incoming funds are invested in long-term projects, but yet instantaneously available to policyholders. Therefore, to allow for even a limited amount of 'casheability', insurance companies <u>must</u> protect themselves by imposing a disintermediation (or market value adjustment) surrender charge. Therefore, from the perspective of the vendor of such products, our model should help determine the appropriate level of restrictions *vis a vis* the promised

[1] An alternative line of reasoning is that in equilibrium, investors should not be compensated for illiquidity, because they can lengthen their trading horizon when faced with such securities. In other words, they can use illiquid instruments to fund long-term liabilities, without demanding any compensation for this inconvenience. Although we do not subscribe to this view, we refer the interested reader to Vayanos and Vila (1999) for a model that pursues this particular approach.

yield.

This so-called liquidity premium cannot, of course, be determined in isolation; it's value will depend on the alternative investments available, and the investor's willingness to make use of them. Thus, we will work in a framework in which there is both a fixed and a variable annuity, and we will impute the investor's level of risk aversion from the allocation chosen between the two annuities. More on this later.

Indeed, there is nascent body of research on the general topic of liquidity, marketability and the bid-ask spread. Various empirical and theoretical studies, such as Silber (1991), Amihud and Mendelson (1991) and more recently, Jacoby, Gottesman and Fowler (2000), Garvey (2001), Brenner, Eldor and Hauser (2001), Dimson and Hanke (2001), Loderer and Lukas (2001), have argued and documented that the yield to maturity, or investment returns, on less liquid financial instruments should be higher compared to their identical liquid counterparts.

However, it appears that limited research has been done on developing a subjective metric for computing the demanded *ex ante* compensation for illiquidity. The exception is a series of papers by Longstaff (1995, 2001). We will provide a more detailed comparison to Longstaff's model, later in our analysis.

The remainder of this paper is organized as follows. In Section 1.1 we demonstrate the simple economic intuition that underlies our model using a basic numerical example. Section 2 develops a formal utility-based model for the liquidity premium in the case of a *savings* (pay-in) annuity, where the time horizon is deterministic and the product is akin to a zero-coupon bond or a Certificate of Deposit. Section 3 solves the model using numerical techniques, with comparative statics provided in Section 3.1 and a comparison to Longstaff's approach discussed in Section 3.2. Then, Section 4 provides a parallel analysis for a *consumption* (pay-out) annuity, where payments are received by the annuitant with embedded longevity insurance. Section 5 concludes the paper.

## 1.1 Numerical Example.

To understand the welfare loss from a lack of liquidity we offer the following example. Consider a hypothetical investor (or policyholder) with \$100,000 to invest. The investor decides to allocate 50% to a fixed annuity (risk free asset) with liquidity restrictions and 50% to a risky equity (variable) annuity. Further, we make the critical assumption that *the investor has picked this allocation because it maximizes his or her expected utility of wealth.*

In the language of Merton (1969), we let $\alpha_t^* = 1/2$, $\forall t \leq T$, denote the optimal allocation to the risky asset, and we let $U_T^*$ denote the maximal expected utility, at the terminal horizon $T$. Merton (1969) demonstrated that an investor with constant relative risk aversion (CRRA) preferences for uncertain wealth at the terminal time $T$, modeled by $u(w) = w^{(1-\gamma)}/(1-\gamma)$, and faced with Geometric Brownian Motion asset dynamics, will select a time-invariant (a.k.a. myopic) investment policy. This well-known Merton result has been generalized to alternative asset processes and consumer preferences. See Kim and Omberg (1996) for more details on the necessary and sufficient conditions for myopic investment policies.

We caution the reader that an $\alpha_t^* = 1/2$ allocation, also known as constant proportional strategy, does *not* imply the portfolio is invested half in equities and half in cash, and then held *as is* until maturity. That is a buy-and-hold strategy and is sub-optimal in a classical Merton framework. Indeed, our 50/50 balance must be maintained by reacting to market movements and rebalancing the portfolio. In other words, rational utility-maximizing behavior requires frequent trading and rebalancing regardless of one's investment horizon or risk preferences. See Browne (1998) for more information on constant proportional strategies.

Suppose, for example, that the general stock market drops 30% within a short period of time. And, as a result, the value of the equity account (a.k.a. variable annuity) drops from $50,000 to $35,000 (= $50,000 x 70%). The investor now has only $85,000 in total, of which, by construction, 41% (= $35,000 / $85,000) is in the equity account, and 59% (= $50,000 / $85,000) is in the fixed annuity. The investor is holding a non-optimal portfolio, which, in theory, should be rebalanced.

A rational investor will want to sell a portion of (or transfer from) the fixed annuity into the equity account to re-establish the optimal 50/50 mix between fixed and variable investments. Specifically, the investor will want to transfer $7,500 from the fixed annuity to the variable account so that $42,500 is invested in fixed assets, and $42,500 is invested in variable assets. Thus maintaining the delicate $\alpha_t^* = 1/2$ mix.

Our main point is that liquidity restrictions in the fixed annuity will impede the optimal process of re-allocation. This is the forgone opportunity cost. Even the prudent buy-and-hold investor will want to rebalance assets after a substantial market movement.

We argue that the only way to make up for the inability to adapt to market movements is to offer an enhanced yield on the fixed annuity. Stated differently, a rational investor will be willing to waive his or her ability to instantaneously rebalance the portfolio in exchange for an enhanced yield on the fixed annuity. Our definition of liquidity yield is meant to provide the same level of economic utility for the constrained investors, as the un-enhanced risk-free asset provides to the unconstrained investor.

TABLE #1 GOES HERE

Table #1 applies our model – which we will fully develop in the next section – to a particular set of parameters and displays our main result. By static utility we mean the maximal utility that can be obtained from picking an asset mix and *holding it* for the entire horizon. By dynamic utility we mean the Merton (1969) values that arise from *rebalancing to maintain a 50/50 mix*. As one can see from the table, *ceteris paribus*, a longer time horizon, lower level of risk aversion and higher subjective growth rate from the market, all imply a larger liquidity premium. The next section presents the formal model that was used to generate Table #1.

# 2 The Utility Model

Our model draws heavily from the classical Merton (1969) framework, and we thus take the liberty of omitting some stages in the derivation. The unrestricted investor can rebalance and allocate assets in continuous time between two assets (a.k.a. sub accounts) under the annuity umbrella. The first is the market (risky, equity) asset that obeys a diffusion process:

$$dV_t = \mu V_t dt + \sigma V_t dB_t \qquad V_0 = 1, \quad 0 \leq t \leq T, \tag{1}$$

where $B_t$ is a standard Brownian motion, $\mu$ is the subjective growth rate of the market, and $\sigma$ is the subjective volatility. This leads to:

$$V_T = e^{(\mu - \sigma^2/2)T + \sigma B_T}. \tag{2}$$

We stress the word *subjective* since the desire to rebalance, and the optimal allocation, will depend critically on the individual's assessment of future market returns and volatility.

The second asset is the Fixed Annuity, or the classically labeled risk-free asset, which obeys:

$$dA_t = rA_t dt, \qquad A_0 = 1 \quad \Longleftrightarrow \quad A_T = e^{rT} \tag{3}$$

In our (simplistic) model, the fixed annuity (bond) pays a constant yield-to-maturity regardless of the time horizon. In practice, of course, one might expect to see a non-flat yield curve, and, as a result, the return on the Fixed Annuity would be a function of the maturity of the product. However, our intention is to exclude, or control for, term-structure premium effects and focus exclusively on liquidity (marketability) issues. As such, we have decided to operate in a flat curve environment. Our main qualitative results are unaffected by the introduction of a stochastic term structure model, which would then force us to keep track of three assets, namely bonds, cash and the variable account.

The end-of-period utility function is of the form:

$$u(w) = \frac{w^{(1-\gamma)}}{1-\gamma}, \qquad \gamma \neq 1, \tag{4}$$

and $u(w) = \ln[w]$ when $\gamma = 1$. Furthermore, without any loss of generality, we assume the investor starts with one (\$1) unit of account (wealth).

Following Merton (1969), the optimal control problem results in a Partial Differential Equation (PDE) which leads to the maximal level of (dynamic) expected utility:

$$EU^*(r|\text{dynamic}) = \frac{1}{1-\gamma} e^{\xi(1-\gamma)T}, \tag{5}$$

where:

$$\xi = r + \frac{(\mu - r)^2}{2\gamma\sigma^2} \tag{6}$$

In this framework,

$$\alpha_t^* = \frac{\mu - r}{\gamma\sigma^2}, \tag{7}$$

which we label the *Merton Optimum*.

In contrast to the dynamic case, a static allocation will induce a maturity-value of wealth which is the linear sum of the monies allocated to the two accounts. The expected utility from this static portfolio – with no liquidity enhancement – is defined as:

$$EU(r|\text{static}) := EU[(1-\alpha)A_T + \alpha V_T] = EU[(1-\alpha)e^{rT} + \alpha e^{(\mu - \sigma^2/2)T + \sigma B_T}]. \tag{8}$$

And, by definition of the optimal allocation:

$$EU^*(r|\text{static}) \leq EU^*(r|\text{dynamic}), \tag{9}$$

with equality occurring when $\alpha^* = 1$ or when $\alpha^* = 0$. We formally define the liquidity premium $\lambda$ as the enhancement to $r$ that will induce the same level of expected utility. In other words:

$$EU^*(r + \lambda|\text{static}) = EU^*(r|\text{dynamic}). \tag{10}$$

In the static case, the maximal expected utility is obtained via:

$$EU^*(r + \lambda|\text{static}) = \max_{\alpha} E\left[\frac{1}{1-\gamma}\left((1-\alpha)e^{(r+\lambda)T} + \alpha e^{(\mu-\sigma^2/2)T+\sigma B_T}\right)^{1-\gamma}\right]. \tag{11}$$

Now, since $B_T$ is normally distributed with mean zero, and variance $T$, equation (11) can be re-written as:

$$EU^*(r + \lambda|\text{static}) = \max_{\alpha} \int_{-\infty}^{\infty} \frac{1}{1-\gamma}\left((1-\alpha)e^{(r+\lambda)T} + \alpha e^{(\mu-\sigma^2/2)T+\sigma\sqrt{T}x}\right)^{1-\gamma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{12}$$

In sum, the (maturity dependent) parameter $\lambda$ is the required yield to compensate for illiquidity. It is an implicit function of the time horizon $T$, the coefficient of relative risk aversion (CRRA) $\gamma$, and the return generating process parameters $r, \mu, \sigma$. Our objective is to solve for $\lambda$.

Of course, any model that attempts to combine risk preferences, $\gamma$, and equity market parameters $\mu, \sigma$, comes face-to-face with the so-called equity risk premium anomaly. A large part of the economics literature is reasonably convinced that $\gamma < 2$. See Feldstein and Ranguelova (2001), or Friend and Blume (1975), for example, for estimates in that range. Likewise, recent work by Mitchell, Poterba, Warshawsky and Brown (1999) in the economic annuities literature, has employed values ranging from $\gamma = 1$ to $\gamma = 3$. Dramatically different evidence is provided by Mankiw and Zeldes (1991) where $\gamma = 35$ and Blake (1996) where $\gamma = 25$. While our methodology is equally applicable in either case – and we provide values for a number of different $\gamma$ values – we prefer to focus our energies and examples on the $\gamma$ in the $(1, 3)$ region, which is consistent with recent papers in the annuity literature. Moreover, a value of $\gamma = 1$, corresponds with log utility, which has intuitively appealing growth optimal properties.

Another thorny issue is that if we use recent (Ibbotson Associates) capital market experience of $\mu - r = 6\%$, and $\sigma = 20\%$, then equation (7) leads to an equity allocation of $\alpha_t^* = 246\%$ for a log-utility investor, and $\alpha_t^* = 123\%$ for a (more risk averse) $\gamma = 2$ investor. Clearly, these allocations are much higher than what is observed in practice. See Campbell (1996), for an in-depth discussion of how to reconcile historical returns and risk preferences.

Thus, to avoid this problem – while at the same time conditioning on a well-balanced portfolio – we decided to invert equation (7) and locate market parameters that 'fit' the Merton model. Specifically, we assume a pre-existing asset allocation $\alpha$, CRRA $\gamma$, and risk premium $\mu - r$, and solve for the (implied) subjective volatility assessment $\sigma = \sqrt{(m-r)/\gamma\alpha}$ that is consistent with Merton's optimum. The implied (subjective) volatility, which is higher than historical values, is motivated by a similar approach in the options market, and attempts to capture the possible model (jump) risks that are not reflected in the classical diffusion approach.

# 3 Solving for the required $\lambda$.

Due to the complexity of equation (11), we are forced to use numerical methods to extract $\lambda$. We start by fixing a value for the risk free rate $r$. Then, for any exogenously imposed value of the CRRA $\gamma$ and subjective rate of return $\mu$, we impute the investor's subjective volatility $\sigma$ from equation (7). For simplicity, consider the case $\gamma \neq 1$ (the case of logarithmic utility $\gamma = 1$ can be treated similarly). Then we are seeking a value of $\lambda$ such that the maximum of

$$F(\alpha, \lambda) = \int_{-\infty}^{\infty} \frac{1}{1-\gamma} \left( (1-\alpha) e^{(r+\lambda)T} + \alpha e^{(\mu - \sigma^2/2)T + \sigma\sqrt{T}x} \right)^{1-\gamma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \qquad (13)$$

equals $U^*(r|\text{dynamic})$. In other words, we are seeking a solution to the pair of equations

$$F(\alpha, \lambda) = U^*(r|\text{dynamic}), \qquad \frac{\partial F}{\partial \alpha}(\alpha, \lambda) = 0. \qquad (14)$$

This may be found numerically using Newton's method, where we alternate Newton steps in the $\lambda$ and $\alpha$ variables. Basically we solve two equations in two unknowns. We start with an initial approximation to the solution ($\alpha_0$,$\lambda_0$). Then, we do a Newton step as a function of the first variable (holding the second variable fixed). This gives a better approximation, denoted by ($\alpha_1$,$\lambda_0$). Then we hold the first variable fixed and look at it as a function of the second variable, and do a Newton step again. This gives an even better approximation $(\alpha_1,\lambda_1)$. Then we go back to the first variable and get a better approximation $(\alpha_2,\lambda_1)$, etc.

To carry this out we require expressions for the functions

$$F, \quad \frac{\partial F}{\partial \alpha}, \quad \frac{\partial^2 F}{\partial \alpha^2}, \quad \frac{\partial F}{\partial \lambda}. \tag{15}$$

But in fact, each of these are easily computed as integrals of simple functions against the standard normal density function. So to carry this out efficiently, all that is required is a method of rapid repeated calculation of such integrals. The method of choice is the Gauss-Hermite integration (see, Press et. al. 1997, Chapter 4), in which a single computation of nodes $x_i$ and weights $w_i$ allows one to write

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x)\, dx \approx \sum_{i=1}^{N} w_i f(x_i) \tag{16}$$

for any regular function $f$. The approximation is exact for polynomials of degree less than $2N-1$. Once again, Table #1 provides values for $\lambda$ for various (subjective) levels of $\mu$ and time horizons $T$.

## 3.1 Comparative Statics.

In Table #2, we display the required liquidity premium, $\lambda$, as a function of the underlying interest rate earned by the risk-free rate. As one can see, the greater the interest rate, the lower is the optimal liquidity premium. Although it might appear from Table #2 that the interest rate, *per se*, is what determines the required yield, it is the *actual spread* between the expected return from the market, $\mu$, and the interest rate that drives this result. Indeed, with a $\mu = 12.5\%$, the higher the level of interest rates, the lower is the spread, and thus the lower is the opportunity cost of *not* being able to rebalance.

TABLE #2 GOES HERE

Once again, we caution the reader that underlying our result is an equity risk premium which also affects the opportunity loss. Thus, for example, when unrestricted cash earns $r = 5\%$, equity is expected to earn $\mu = 12.5\%$, the investor has a CRRA $\gamma = 2$, <u>and</u> a pre-existing portfolio of 50% cash and 50% equity, the implied subjective volatility assumption is $\sigma = 31.62\%$.

Finally, in Table #3, we display the required liquidity premium as a function of the pre-existing asset allocation. Thus, for example, a $\gamma = 3$ individual with a desired 20% allocation to risky equity, and an 80% allocation to unrestricted cash, will demand a liquidity premium of 35.86 basis points per annum as compensation for being unable to trade during a 10-year investment horizon.

TABLE #3 GOES HERE

As one can see from Table #3, the relationship between desired (or pre-existing) equity holdings, and the demanded liquidity premium resembles an inverted parabola, and is zero at both ends. The intuition is as follows. A rational individual with a desired (or pre-existing) allocation of either, 0% or 100% unrestricted cash, will not be engage in any trading during the length of the investment horizon (with probability one) since there will never be a need to rebalance. However, as the portfolio moves towards a more balanced composition, the probability and magnitude of rebalancing increases, thus magnifying the required liquidity premium for not being able to trade.

The same inverted-parabolic relationship exists for higher levels of risk aversion, but in a decreasing manner since *the opportunity cost of not being able to trade is lower.* For example, despite much economic evidence to the contrary, one sometimes finds relatively high values of the risk aversion parameter $\gamma$ in use. For example, Blake (1996) has estimated CRRA values as high as $\gamma = 25$ in the UK. Thus, if we were to take $\gamma = 25$, for example, then the only way a dynamic allocation of $\alpha = 50\%$ could be rational would be if the individual's estimate of future volatility was particularly low, in which case there is less difference between the risky asset and the risk-free one, so the liquidity premium should also be low. This is indeed the case. To get a sense of the magnitudes, if $\mu = 20\%$ and $T = 10$, we compute $\alpha = 49.74\%$ and $\lambda = 7.53$ b.p., but at the same time find that $\sigma = 10.95\%$

## 3.2 Comparison to Longstaff's Model.

While we took a similar approach to computing the welfare loss for liquidity restrictions, our paper differs from Longstaff's 2001 work in a number of substantial ways. First, our model assumed a general constant relative risk aversion (CRRA) utility specification, in contrast to Longstaff's logarithmic utility model. This allowed us to explore the critical impact of risk aversion on the value of liquidity, as well as the effect of holding period restrictions. Indeed, as Tables 1-3 indicated, risk aversion played an important role in determining the required liquidity premium. *Ceteris paribus*, the greater the aversion to risk, the lower is the required liquidity premium. Also, while Longstaff modeled illiquidity in a stochastic volatility environment, and used trading strategies that were of bounded variation, we operated in a much simpler Merton (1969) environment, which allowed for closed-form solutions to the optimal portfolio holdings. (We traded-off stochastic volatility for general utility.) Our liquidity premium – which was formulated as a yield, as opposed to Longstaff's discount – was obtained by solving a one dimensional integral equation.

However, the most important distinction with Longstaff (2001), was that we focused on the individual's pre-existing portfolio and asset allocation as a determinant of the liquidity premium. As one can see from Table #3, an individual with a very low, or very high, level of holdings in the illiquid bond (annuity) would not require as much compensation as the individual with a relatively well-balanced portfolio. The liquidity premium is directly related to the probability (and magnitude) of having to trade and rebalance during the life of the restriction. If the pre-existing optimal portfolio is well-balanced – i.e. close to equal amounts of equity and cash – there is a higher chance of the portfolio falling out of balance, and thus requiring trading to move back to the optimum.

As such, our conclusions complement Longstaff (2001), in that we concur that "discounts for illiquidity can be substantial", but we also demonstrate that the magnitude depends on the individual's risk aversion and pre-existing portfolio.

# 4 Application to Payout Annuities and Longevity Insurance

Within the universe of annuities, the most natural context for the foregoing section is the accumulation phase, during which contributions are held and invested prior to retirement. We turn now to the payout phase of the life annuity. For simplicity we assume that this involves the purchase of an immediate life annuity at time $t = 0$, entitling the holder to a continuous stream of payments, terminating upon death, which is now a random time $t = T$.

The annuity can be some combination of a fixed immediate annuity (FIA), which provides a fixed payment per unit time, and a variable immediate annuity (VIA) which provides a payment per unit time that varies depending on the value of some market asset $V_t$. If $w$ dollars of the FIA are purchased, the consumer is entitled to continuous payment stream of $C_t^F = w/a_x(r)$ dollars per unit time, where the unit price of the FIA is:

$$a_x(r) = \int_0^\infty e^{-rt}({}_tp_x)dt. \tag{17}$$

Here $r$ denotes the risk-free interest rate, and $({}_tp_x)$ is the probability that the individual will survive to time $t$, conditional on being alive at the annuity purchase age $x$. The normalization is that each unit of the FIA pays \$1 per unit time.

Likewise, if $w$ dollars of the VIA are purchased, the consumer receives payments based on $w/a_x(h)$ units of the market asset per unit time, where $h$ is the assumed interest rate (AIR). In other words, at time $t$, payments accumulate at the rate of $C_t^V = we^{-ht}V_t/a_x(h)$ dollars per unit time, where we have normalized the market asset so that $V_0 = 1$.

As before, we will compare liquid and illiquid annuities. In the liquid case, the consumer is free to exchange FIA units for an economically equivalent number of VIA units at any time, and vice versa. In the illiquid case, the number of FIA and VIA units is fixed at the time of purchase. Other things being equal, the liquid annuity would provide greater utility to the consumer, so to compensate for this the illiquid annuity must provide an enhanced rate of return. As in the preceding section, we assume that it is the FIA that is so enhanced. In this context, we take this to mean that an investment of $w$ dollars in the FIA produces a payment stream of $C_t^F = w/a_x(r+\lambda)$ dollars per unit time, where $\lambda$ is the demanded liquidity premium.

We will assume that the AIR is chosen to be the risk-free rate, so that $h = r$. Such a restriction is not uncommon in annuity products available for sale, and is in fact typical of the liquid ones. Illiquid annuities more commonly allow their purchasers to choose a value of the AIR, but all such choices are deemed to be economically equivalent. Since our principal interest is in the liquidity premium, we will require $h = r$ for both liquid and illiquid annuities, so that the effect of liquidity is not confounded with that of a flexible AIR.

In the preceding section the consumer's utility involved only end-of-period wealth, since there were no funds available for consumption prior to that time horizon. In the present case, it is exactly the utility of consumption that is of interest, discounted to take account of the time-value of money. Thus, if $C_t$ denotes the payment stream generated by the life annuity, and if the function $u(.)$ denotes the consumer's personal utility of consumption, then the mix between the fixed and variable annuities will be selected so as to maximize:

$$E\left[\int_0^T e^{-rt}u(C_t)dt\right] = \int_0^\infty e^{-rt}({}_tp_x)E[u(C_t)]dt. \tag{18}$$

Of course we assume independence of asset returns and mortality.

As before, we will assume an optimal 50/50 mix between the fixed and variable annuities in the liquid case, and then impute model parameters. We continue to assume geometric Brownian dynamics for the risky asset $V_t$, so

$$dV_t = \mu V_t dt + \sigma V_t dB_t, \qquad V_0 = 1 \tag{19}$$

In the previous section, we postulated several values for the risk aversion (CRRA) parameter $\gamma$ and then imputed the individual's forecast of the future volatility $\sigma$ that would be consistent with a balanced portfolio allocation. To illustrate the flexibility of our approach, in this section we will start at the opposite end. Instead, we use a volatility based on historical capital market data, namely $\sigma = 20\%$, and we then impute a corresponding level of risk aversion, assuming a balanced portfolio allocation. We also take historical levels of $\mu = 11\%$ and $r = 5\%$. These numbers are consistent with the widely quoted Ibbotson Associate numbers used by practitioners.

Charupat and Milevsky (2002) consider the asset allocation problem in the setting of liquid annuities. Assuming $h = r$, and an exponential or Gompertz mortality function, they show that the Merton optimum

$$\alpha^* = \frac{\mu - r}{\gamma\sigma^2} \tag{20}$$

remains optimal in this new setting. In fact this can be proved more generally – and is actually alluded to in Chapter 18 of Merton (1994) – and does not depend on the parametric form of the survival probabilities $({}_tp_x)$. Denote by $\phi_t$ and $\psi_t$ the number of units of the FIA and VIA held at time $t$, and assume that $h = r$. Then the payment stream is $C_t = \phi_t + \psi e^{-rt}V_t$, and it can be shown that the optimal choice of $\phi_t$ and $\psi_t$ obeys:

$$\alpha^* = \frac{\psi_t e^{-rt}V_t}{\phi_t + \psi_t e^{-rt}V_t} = 1 - \frac{\phi_t}{\phi_t + \psi_t e^{-rt}V_t} \tag{21}$$

for $\alpha^*$as above. In particular, from the assumption that $\alpha^* = 50\%$, and the given (Ibbotson Associates) values for $\mu, r$ and $\sigma$, we may impute a CRRA value of $\gamma = 3$, regardless of the form of the conditional probability of survival $({}_tp_x)$. In other words, if the individual has a coefficient of relative risk aversion of $\gamma = 3$, and is faced with a market in which the expected return from the risky asset is $\mu = 11\%$, with a volatility of $\sigma = 20\%$, when the risk free rate is $r = 5\%$, then he/she will allocate exactly $\alpha = 50\%$ to each of the two asset classes.

It can further be shown that with this choice of allocation,

$$E[u(C_t)] = \frac{1}{1-\gamma} e^{\beta t} \left( \frac{w}{a_x(r)} \right)^{1-\gamma}, \tag{22}$$

where

$$\beta = \frac{(1-\gamma)(\mu - r)^2}{2\gamma\sigma^2} \tag{23}$$

Thus:

$$U^* = \int_0^\infty e^{-rt} ({}_tp_x) E[u(C_t)] dt = \frac{w^{(1-\gamma)}}{1-\gamma} \frac{a_x(r-\beta)}{a_x(r)^{(1-\gamma)}}, \tag{24}$$

in the dynamic liquid case.

In the static (illiquid) case, an initial allocation of $\alpha$ to the risky asset will result in holding $\phi_t = (1-\alpha)w/a_x(r+\lambda)$ FIA units, and $\psi_t = \alpha w / a_x(r)$ VIA units, and in a utility:

$$\begin{aligned} F(\alpha, \lambda) &= \int_0^\infty e^{-rt} ({}_tp_x) E[u(C_t)] dt \\ &= \int_0^\infty e^{-rt} ({}_tp_x) \int_{-\infty}^\infty \frac{w^{1-\gamma}}{1-\gamma} \left( \frac{1-\alpha}{a_x(r+\lambda)} + \frac{\alpha e^{(\mu - r - \sigma^2/2)t + \sigma\sqrt{t}z}}{a_x(r)} \right)^{1-\gamma} \\ &\times \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz dt \end{aligned} \tag{25}$$

Our goal is, as before, to find the liquidity premium $\lambda$ such that maximizing $F(\alpha, \lambda)$ over $\alpha$ reproduces the dynamic utility $U^*(r|\text{dynamic})$. We will do so assuming Gompertz mortality, corresponding to an exponentially increasing hazard rate (force of mortality) of the form:

$$h_{x+t} = \frac{1}{b} e^{(x+t-m)/b}, \tag{26}$$

$(m, b)$ are the Gompertz parameters and $x$ is the individual's age at the time of purchase. In this case the survival probability takes the form:

$$({}_tp_x) = \exp\left( b h_x (1 - e^{t/b}) \right). \tag{27}$$

As before, we use Newton's method to carry out the maximization and root finding, and we use Gauss-Hermite quadrature to rapidly evaluate the Gaussian integral in the expression for $F(\alpha, \lambda)$. We carry out the time integral using a related method, namely Gauss-Laguerre quadrature.

The Gauss Laguerre nodes and weights are optimized for computing integrals of the form

$$\int_0^\infty e^{-at} t^c f(t) dt \tag{28}$$

where $f$ is well approximated by a polynomial function. We use $c = 0$ and must be careful to choose $a$ in a narrow range of values for which the method is stable when applied to our integrands. But having done so, this gives a rapid and accurate algorithm.

Recall that our market parameters were $\mu = 11\%$, $r = 5\%$ and $\sigma = 20\%$, and from the dynamic allocation of 50/50 to the FIA and VIA we imputed $\gamma = 3$. We consider two cases, both corresponding to an age of 62 years at the time of annuitization. The first uses Gompertz parameters fit to the U.S. Society of Actuaries female (IAM1996) mortality data (namely $b = 8.78$ and $m = 92.63$) and yields an optimal allocation of $\alpha = 48.40\%$ and a liquidity premium of $\lambda = 13.07$ basis points. In the second case we use male mortality parameters $b = 10.5$ and $m = 88.18$, and compute $\alpha = 48.50$ and $\lambda = 12.52$ b.p.

To understand the factors influencing these results, one can calculate the conditional life expectancy, resulting in figures of $e_{62} = 26.62$ years (female) and $e_{62} = 22.78$ years (male). Using those time-horizons in the fixed maturity problem of the previous section gives $\alpha = 47.68\%$, $\lambda = 24.22$ b.p., and $\alpha = 47.93\%$, $\lambda = 21.87$ b.p. respectively. These premiums are substantially higher than the Gompertz figures just computed, and a moment's reflection will spot the reason why. We saw that the liquidity premium increases rapidly with the time horizon, and in the annuity context most of the payments occur significantly earlier than the lifetime itself. Thus the effect of spreading payments out over the residual lifetime should be to reduce the liquidity premium. Indeed, even if the residual lifetime were to take on a deterministic value $T$, the mean time a payment is received is

$$\frac{1}{T}\int_0^T t dt = \frac{T}{2}. \tag{29}$$

Thus to appreciate the sensitivity of the results to the randomness of the life horizon $T$, we should not compare with the results of the preceding section, but rather with other lifetime distributions having the same means. As an extreme case, we compare the Gompertz results with deterministic lifetime distributions, that is, with survival functions

$$({}_tp_x) = \begin{cases} 1, & t < t_0 \\ 0, & t \geq t_0 \end{cases} \tag{30}$$

where $t$ is set to the mean residual Gompertz lifetimes. Because of the discontinuity in $({}_tp_x)$ we use yet another quadrature method (Gauss-Legendre this time) for the $t$ integral. This gives optimal allocations of $\alpha = 48.48\%$ (female) and $\alpha = 48.63\%$ (male), and liquidity premiums of $\lambda = 13.02$ b.p (female) and $\lambda = 12.20$. These premiums are extremely close to those obtained under Gompertz mortality, which suggests that the premiums are not highly sensitive to the precise form of the hazard rate. Note however that both Gompertz figures are slightly higher, and it is tempting to describe the difference as a small additional premium for mortality risk.

# 5 Conclusion.

This paper has argued that the value of liquidity in an annuity product, or the lack thereof, can be assessed by returning to first principles. We did this by locating the required enhancement to the risk-free rate which compensated for the inability to rebalance an investment portfolio. Our main mathematical problem was to locate the yield, which we denoted by $\lambda$, that equated maximal utility in a static portfolio to the (greater) utility from a portfolio that could be dynamically rebalanced in a Merton framework.

Using our model with recent capital market parameters, we argue that a log-utility ($\gamma = 1$) investor, with a pre-existing 50/50 asset mix between fixed and variable *savings* annuities, would demand a liquidity premium of between 45 - 145 basis points per annum as compensation for the inability to rebalance during a 10-year period.

However, for investors that are more risk averse ($\gamma > 1$), and/or who are faced with shorter liquidity restrictions, the compensating premium is lower. Indeed, for a 1-year period, and coefficient of relative risk aversion ($\gamma = 3$) the premium ranges from *only* 2 - 8 basis points per annum above the risk-free yield. Likewise, as Table #3 indicated, the pre-existing asset allocation has a dramatic impact on the liquidity premium as well. For the above mentioned log-utility investor, the premium ranged from 0 to 85 basis points, depending on current portfolio holdings. As such, we are careful to conclude that the question of liquidity is personal in nature, since it depends on attitudes towards financial risk and subjective expectations about future investment returns. Thus, there is no universal compensation for trading restrictions.

However, regardless of the magnitude of this effect, our paper supports the argument that impeding the consumers ability to continuously rebalance his or her investment portfolio is detrimental to economic utility and financial wealth. This is regardless of their pre-existing asset allocation, investment time horizon or subjective market expectations.

Finally, research currently underway by the authors will go towards developing a model in which only one-sided trading restrictions are imposed so that additional assets can be purchased, but not sold. We anticipate that the liquidity premium will be lower in this case, but the amount by which it is reduced remains an open question.

# 6 References


1. Amihud, Y. and Mendelson, H. (1991), "Liquidity, Maturity and the Yields on U.S. Treasury Securities", *Journal of Finance*, Vol. 46, pg. 1411-1425.

2. Blake, D., Burrows W. (2001), "Survivor Bonds: Helping to Hedge Mortality Risk", *The Journal of Risk and Insurance*, Vol. 68(2), pg. 339-348.

3. Blake, D. (1996), "Efficiency, Risk Aversion & Portfolio Insurance: An Analysis of Financial Asset Portfolios Held by Investors in the UK", *Economic Journal*, Vol. 106, pg. 1175-1192.

4. Brenner, M., Eldor, R. and Hauser S. (2001), "The Price of Options Illiquidity", *The Journal of Finance*, Vol. 56(2), pg. 789-805.

5. Brown, J. R., Poterba, J. M. (2000), "Joint Life Annuities and Annuity Demand by Married Couples", *The Journal of Risk and Insurance*, Vol. 67(4), pg. 527-554.

6. Browne, S. (1998), "The Return on Investment from Proportional Portfolio Strategies", *Advances in Applied Probability*, 30(1), pg. 216-238.

7. Charupat, C. and Milevsky M.A., (2002) "Asset Allocation in Variable Annuities: A Note", *Insurance: Mathematics and Economics*, Vol. 30(2), pg. 199-210.

8. Campbell, J. Y. (1996), "Understanding Risk and Return", *The Journal of Political Economy*, Vol. 104(2), pg. 298-344.

9. Dimson, E. and Hanke, B. (2001), "The Expected Illiquidity Premium: Evidence from Equity Index-Linked Bonds", *Working Paper*, London Business School.

10. Feldstein, M. and Ranguelova E., (2001), "Individual risk in an investment-based social security system", *American Economic Review*, 91(4), pg. 1116-1125.

11. Friend, I. and Blume, M.E. (1975), "The Demand for Risky Assets", *American Economic Review*, Vol. 65, pg. 900-922.

12. Garvey, G.T. (2001), "What is an Acceptable Rate of Return for an Undiversified Investor?", *Working Paper*, Claremont Graduate School.

13. Ibbotson Associates (2000), *Stocks, Bonds, Bills and Inflation: 2000 Yearbook*, Chicago, IL.

14. Jacoby, G., Gottesman A. and Fowler, D. (2000) "The Capital Asset Pricing Model and the Liquidity Effect: A Theoretical Approach", *Journal of Financial Markets*, 3(1), pg. 69-81.

15. Kim, T.S. and Omberg, E. (1996), "Dynamic Nonmyopic Portfolio Behavior", *Review of Financial Studies*, Vol. 9(1), pg. 141-161.

16. Loderer, C. and Lukas, R. (2001), "The Pricing Discount for Limited Liquidity: Evidence from Switzerland", *Working Paper*, University of Bern, Switzerland.

17. Longstaff, F.A. (1995), "How Much Can Marketability Affect Security Values?", *The Journal of Finance*, Vol. 50(5), pg. 1767-1774.

18. Longstaff, F.A. (2001), "Optimal Portfolio Choice and the Valuation of Illiquid Securities", *The Review of Financial Studies*, Vol. 14(2), pp. 407-431.

19. Mankiw, N. and Zeldes, S. (1991), "The Consumption of Stockholders & Nonstockholders", *The Journal of Financial Economics,* Vol. 29, pg. 97-112

20. Merton, R.C. (1969), "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case", *Review of Economics and Statistics*, Vol. 51, pg. 247-257.

21. Merton, R.C. (1994), *Continuous-Time Finance* (Revised Edition), Blackwell Publishers, Oxford, UK.

22. Mitchel O., J. Poterba, M. Warshawsky and J. Brown, (1999) "New Evidence on the Money's Worth of Individual Annuities", *American Economic Review*, Vol. 89(5), pg. 1299-1318.

23. Press, W. H., Teukolsky, S.A., Vetterling, and W.T., Flannery, B.P. (1997), *Numerical Recipes in C (2nd Ed.)*, Cambridge University Press.

24. Silber, W. (1991), "Discounts on Restricted Stock: The Impact of Illiquidity on Stock Prices", *Financial Analysts Journal*, Vol. 47, pg. 60-64.

25. Stapleton, R.C. and Subrahmanyam, M.G. (1979), "Marketability of Assets and the Price of Risk", *Journal of Financial and Quantitative Analysis*, Vol. 14, pg. 1-10.

26. Vayanos, D., and Vila, J. (1999), "Equilibrium Interest Rate and Liquidity Premium with Transaction Costs", *Economic Theory*, Vol. 13, pg. 509-539.

**Table #1: Required Liquidity Premium Spread, in basis points, per annum.**
*Assuming a desired 50/50 allocation to (risky) equities and (safe) cash earning 5% p.a.*

$$\boxed{U(w) = w^{(1-g)} / (1-g)}$$

Coefficient of Relative Risk Aversion

| T = 15 Years | $\gamma = 1$ | $\gamma = 2$ | $\gamma = 3$ |
|---|---|---|---|
| $\mu$ = 12.5% | 62.29 | 34.59 | 23.92 |
| $\mu$ = 15.0% | 98.66 | 55.73 | 38.76 |
| $\mu$ = 17.5% | 139.34 | 79.80 | 55.75 |
| $\mu$ = 20.0% | 183.36 | 106.24 | 74.48 |

| T = 10 Years | $\gamma = 1$ | $\gamma = 2$ | $\gamma = 3$ |
|---|---|---|---|
| $\mu$ = 12.5% | 47.56 | 25.84 | 17.74 |
| $\mu$ = 15.0% | 77.04 | 42.50 | 29.33 |
| $\mu$ = 17.5% | 110.78 | 61.90 | 42.90 |
| $\mu$ = 20.0% | 147.98 | 83.59 | 58.14 |

Coefficient of Relative Risk Aversion

| T = 5 Years | $\gamma = 1$ | $\gamma = 2$ | $\gamma = 3$ |
|---|---|---|---|
| $\mu$ = 12.5% | 28.10 | 14.80 | 10.05 |
| $\mu$ = 15.0% | 47.04 | 25.07 | 17.09 |
| $\mu$ = 17.5% | 69.54 | 37.44 | 25.62 |
| $\mu$ = 20.0% | 95.12 | 51.68 | 35.47 |

| T = 1 Year | $\gamma = 1$ | $\gamma = 2$ | $\gamma = 3$ |
|---|---|---|---|
| $\mu$ = 12.5% | 6.67 | 3.38 | 2.27 |
| $\mu$ = 15.0% | 11.67 | 5.94 | 3.98 |
| $\mu$ = 17.5% | 17.95 | 9.17 | 6.16 |
| $\mu$ = 20.0% | 25.45 | 13.05 | 8.78 |

*Notes: For example, an investor with a 50/50 allocation to equities and cash, with CRRA = 1 (a.k.a. log utility) preferences, would require a yield enhancement (lambda) of 77 basis points on the cash account (I.e. 5.77%) to compensate for the inability to rebalance for 10 years; this is assuming they expected the equity account to earn 15% p.a. during this period.*
*In contrast, if the investor expects to earn 17.5% from the equity account, they would require a 111 basis point liquidity spread.*
*Likewise, for a fixed 50/50 allocation and the same (subjective) equity return, a higher CRRA (I.e. a more risk averse investor), requires less compensation for the inability to rebalance*

**Table #2**

*Liquidity premium as a function of the risk-free interest rate,*
*conditional on a 50/50 allocation to Equities vs. Cash,*
*and assuming an equity return expectation of 12.5%,*
*over a 10-year investment horizon.*

| **Interest** | CRRA = 1 | CRRA = 2 | CRRA = 3 |
|---|---|---|---|
| 4% | 58.77 | 32.14 | 22.11 |
| 5% | 47.56 | 25.84 | 17.73 |
| 6% | 37.21 | 20.07 | 13.74 |
| 7% | 27.81 | 14.89 | 10.16 |

*See Notes to Table #1*

**Table #3**

*Liquidity premium as a function of asset allocation,*
*Equities (risky asset) vs. Cash (safe asset).*
*Assuming an equity return expectation of 15%,*
*cash earning 5%, and a 10-year investment horizon.*

| **Equity %** | CRRA = 1 | CRRA = 2 | CRRA = 3 |
|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 |
| 10 | 47.92 | 39.00 | 31.44 |
| 20 | 74.12 | 48.74 | 35.86 |
| 30 | 82.38 | 49.45 | 35.18 |
| 40 | 82.11 | 46.87 | 32.74 |
| 50 | 77.04 | 42.50 | 29.33 |
| 60 | 68.60 | 36.83 | 25.16 |
| 70 | 57.29 | 29.97 | 20.28 |
| 80 | 42.96 | 21.84 | 14.63 |
| 90 | 24.79 | 12.11 | 8.00 |
| 100 | 0.00 | 0.00 | 0.00 |

*See Notes to Table #1*